Incorporating individual heterogeneity into mark-recapture models

by

Jessica Helen Ford

BSc (Mathematics, Neuroscience), MSc (Medical Statistics)

Submitted in fulfillment of the requirements for the Degree of

Doctor of Philosophy in Quantitative Marine Science

(A joint CSIRO and UTAS PhD program in quantitative marine science)

University of Tasmania

March 2013

Statement of Declaration

I declare that this thesis contains no material which has been accepted for a degree or diploma by the University or any other institution, except by way of background information and duly acknowledged in the thesis, and to the best of my knowledge and belief no material previously published or written by another person except where due acknowledgment is made in the text of the thesis, nor does the thesis contain any material that infringes copyright.

This thesis may be made available for loan and limited copying in accordance with the *Copyright Act of 1968*

Jessica Helen Ford

Dedication

To my parents, for much love, support, and encouragement throughout life.

Abstract

Mark-recapture analysis is a fundamental tool for understanding populations, since it allows the estimation of demographic parameters, such as survival, movement and reproduction, which can be used to infer population status and predict dynamics. As individuals in wild populations do not all behave in the same way, a challenge is presented in the collection and analysis of these data. Within a natural population, animals may exhibit substantial individual variation which can manifest through these demographic parameters. Inherent individual differences in movement and behavior can introduce bias into mark-recapture estimates (most notoriously, of population size), and are often of considerable interest in their own right.

There has been much focus in mark-recapture research on the development of methods to account for individual heterogeneity, yet easily applied, accurate methods are still lacking. The most natural, but computationally complex, approach for modeling individual heterogeneity assumes a continuous distribution using random effects. This method introduces the complexity of solving for the individual random effects which has been a stumbling block of much work in the mark-recapture field. The focus of this thesis is the development of methods to better estimate individual heterogeneity in mark-recapture data.

In chapter 1 I introduce the concepts arising in this thesis and briefly outline techniques for modeling individual heterogeneity. Chapter 2 explores the population consequences of individual heterogeneity in spatial use in the context of a marine protected area. Using population projections, I explore the population consequences of individual heterogeneity in proportion of time spent inside a marine protected area. The projections indicate that individual heterogeneity in spatial use and site fidelity could have important implications under certain conditions for the dynamics of populations managed using marine protected areas. In several scenarios, high individual heterogeneity resulted in larger population size and positive population trajectories, compared to decline and eventual extinction with low individual heterogeneity in site fidelity.

I then present three novel statistical approaches for handling individual heterogeneity using random effects. The first, developed in chapter 3, is an approach using Laplace approximation with Gaussian random effects, implemented in the language Automatic Differentiation Model Builder (ADMB). In chapter 4 I develop a Markov chain Monte Carlo sampling framework with a parametric distribution for the individual heterogeneity. This is extended in chapter 5 to incorporate the non-parametric Dirichlet process prior. The natural subgroups often seen in mark-recapture studies, and the complexity of real mark-recapture data means that parametric and discrete style models can be insufficient. Non-parametric models avoid these often restrictive assumptions. The Dirichlet process prior is a flexible extension to a parametric model as it avoids assumptions about the functional form of the distribution, and it extends discrete style models to the infinite limit by avoiding any prespecifications about the number of groups. Each of these methods was tested using simulated data. In each case the simulation studies demon-

strated accurate estimation of true parameter values with random effects. In the case of the Dirichlet process, the simulation studies were used to explore the ability and limits of the Dirichlet process in identifying multiple behavioural modes.

The methods are applied to data for up to 1100 individually identified North Atlantic humpback whales, where an unseen individual may be present but not seen, temporarily absent, or dead. There was some evidence of individual heterogeneity in site fidelity and multimodality in the probability of observation.

Contents

List of Figures

List of Tables

Acknowledgments

Needless to say, the past three and a half years would not have been possible without the help and support of many people.

Firstly I would like to thank my supervisors. I have learnt much off all of you, and received the best support and guidance throughout my candidature. Chris, many thanks for your knowledge, guidance and support, from when we first started with the snails. Toby, you've provided daily guidance, knowledge, support, and discussions throughout the past few years which have been invaluable and enjoyable, and for which I am incredibly grateful. Mark B, many thanks for much guidance, knowledge, and for patience in discussions when I couldn't see the light and lacked clarity and finesse in getting to the point (most discussions?). Mark H, many thanks for your help and guidance when sought.

I was supported throughout my candidature by the Quantitative Marine Science (QMS) program run in partnership between CSIRO and UTAS, by the CERF marine biodiversity hub (now NERP) and by CSIRO. Many thanks to those running the QMS program, particularly Denbeigh, Heidi, Tom and Spoon for much help. Many people at CSIRO have provided help, in particular I would like to thank three: Peter Campbell who provided help with Condor and Perl scripts which made running a ridiculous number of ADMB simulations manageable; Paavo Jumppanen for help with C++;

and to Rich Hillary for welcoming many impromptu corridor and doorway discussions, especially during my getting to grips with the delights of MCMC and Dirichlet processes.

Jooke Robbins and the PCCS crew very generously gave me access to their valuable data. The work they have done (and continue to do) is amazing, and I would like to thank them for their collaboration.

Thanks to Annette Dobson who gave me the opportunity to start on this road, and supported and mentored me many years ago.

Special thanks to my fellow students Mel, Marie, Celeste, Jim and Kilian, it's been a pleasure. Marie, thanks for friendship, many morning coffees and delightful discussions. Mel, you've taught me a lot. I've thoroughly enjoyed our many discussions, much (welcome) distraction and a fabulous friendship. The process has been so much better for having you two around, thanks for the laughs.

Finally, I would like to thank my family and friends, for lots of fun times. I managed to keep the work-life balance well balanced thanks to you.

Ben, thank you for love and encouragement through the last couple of years of the typo, and for much fun and time in life away from the screen.

Mum and Dad, thank you for love, support and encouragement throughout life, in all matters, always.

Chapter 1

Introduction

1.1 Overview

Studying populations of wild animals using mark-recapture experiments involves sampling from the population and marking the sampled animals (with a tag, band or through unique natural markings) and then returning them to the population. On subsequent sampling occasions, the marked animals may be re-captured or identified, along with marking of new animals. Mark-recapture analysis is a fundamental tool for understanding populations, since it allows the estimation of demographic parameters, such as survival, movement and reproduction, which can be used to infer population status and predict dynamics. However, one challenge in the collection and analysis of this data is that individuals in wild populations do not all behave in the same way. Within a natural population, animals may exhibit substantial individual variation in a range of parameters.

Individual heterogeneity within species can include factors such as prey preference, foraging techniques (Bolnick et al., 2003) and inherent differences in behavior (Hammond, 1990; McNamara and Houston, 1996). In

some cases, some measurable covariate such as age, length or sex may be able to explain the individual differences. However, it is often impossible to collect this information for all animals. Whether we have access to this extra information or not, in most cases there will still be latent individual heterogeneity which cannot be explained by available covariates. Failing to account for this latent individual variation can result in incorrect inference and give rise to biased estimates of demographic parameters of interest (Seber, 1982). Most notably, if unaccounted for, capture heterogeneity is known to bias abundance estimates. For example, under-estimation in abundance may occur if some animals have higher capture probabilities. Methods to effectively estimate this latent individual variation (i.e. not captured through covariates) are thus of statistical importance, but the the variation is also of interest in its own right. The focus of this thesis is the development of statistical methods to better model latent individual heterogeneity arising from mark-recapture data.

Along with demographic parameters such as survival and reproduction, individual variation in spatial use within natural populations is common. Implementation of spatial protection for many species is often in the form of marine protected areas (MPAs). A challenge for MPAs is striking a balance between an area large enough to be biologically relevant whilst also small enough to be politically viable (Ashe et al., 2010). Heterogeneity in spatial use within a population may affect the ability of the reserve to protect the population both inside and outside the reserve area. The need for accurate quantitative understanding of the effectiveness of MPAs is apparent (Hooker et al., 2011; Gormley et al., 2012), but to date there are minimal tools for achieving this. As a motivating example, I show in chapter 2 how individual variation can affect the performance of a MPA. Following this, the remain-

ing chapters focus on the development of statistical methods to estimate individual heterogeneity.

1.2 North Atlantic Humpback Whales

The methods developed in chapters 3, 4 and 5 in this thesis estimate latent individual heterogeneity in detection and movement parameters. They are applied to a mark-resight data set on a subpopulation of North Atlantic humpback whales sighted in the Stellwagen Bank National Marine Sanctuary (SBNMS), in the Gulf of Maine (Figure 1.1). Researchers from the Provincetown Centre for Coastal Studies began documenting North Atlantic humpback whales in the Gulf of Maine in 1975 and have to date individually identified over 1200.

1.2.1 Background

Humpback whales (*Megaptera novaeangliae*) are distributed worldwide, with summer feeding ranges in mid to high-latitudes and winter breeding in low-latitude areas (Clapham and Mead, 1999). They can be uniquely identified by their natural markings: through the shape of their flukes and through patterns from natural pigmentation (Hammond, 1986).

Two large ocean-wide catalogues exist for North Atlantic humpback whales: the North Atlantic Humpback Whale Catalogue (NAHWC) and the Years of the North Atlantic Humpback Whales (YONAH). Individual identification of whales within the NAHWC began in 1968 and presently contains photographs of over 5000 individuals (Barco et al., 2002). YONAH was an ocean-basin-wide collaborative study, with the majority of photographs from a focused effort during 1992/93 (Barco et al., 2002).

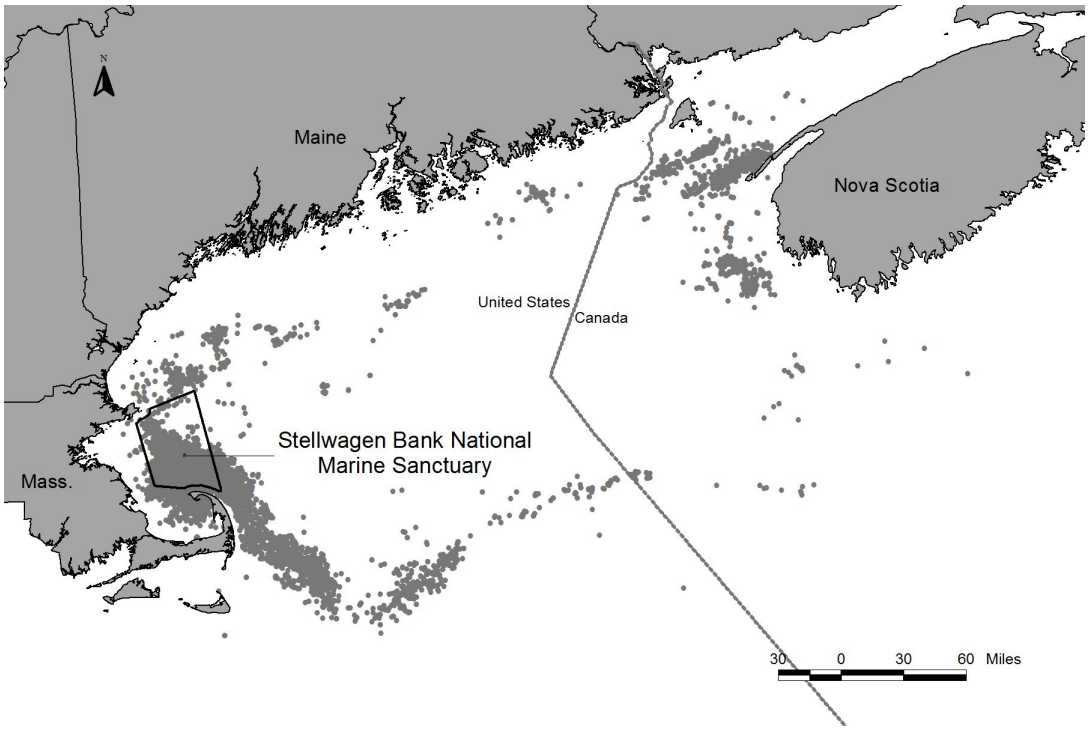

Figure 1.1: Stellwagen Bank National Marine Sanctuary (study site) in relation to the greater Gulf of Maine humpback whale feeding ground. Grey circles represent sightings of individually identified humpback whales made by the Provincetown Center for Coastal Studies, 1979-2005. Sightings are not corrected for effort.

The majority of the North Atlantic humpback whales breed over winter in the West Indies; a small number are thought to use the breeding grounds around the Cape Verde Islands (Stevick et al., 1998). During summer, the whales disperse to six summer feeding regions. Although historically treated as a single stock, the six summer feeding regions in the North Atlantic hold relatively discrete subpopulations (Clapham and Mayo, 1987), with individuals demonstrating strong site fidelity to a particular feeding region over many years. Feeding sites include the Gulf of Maine, eastern Canada, west Greenland, Iceland and Norway (Katona and Beard, 1990), and patterns of movement suggest perhaps four distinct subpopulations (Stevick et al., 2006). Individual humpback whales show high maternally directed site fidelity to these summer feeding ranges, as calves follow their mothers from breeding to feeding grounds (Clapham and Mayo, 1987).

1.2.2 Gulf of Maine Catalogue

The Gulf of Maine is the southern most summer feeding ground for the North Atlantic humpback whales. Individual humpback whales have been intensively studied in this region since the late 1970s. The SBNMS (Figure 1.1) is one of several important feeding sites for North Atlantic humpback whales which summer in the Gulf of Maine. Due to the consistent aggregation of humpback whales and other marine life, the SBNMS was nominated as a national sanctuary in 1992. This area is not only an important feeding ground for the North Atlantic humpback whales, but is also a busy recreation and transportation area for humans with high levels of commercial and recreational vessel traffic. This overlap has resulted in many injuries to the whales from ship collision and entanglement in fishing gear (Robbins and Mattila, 2004). Although both commercial and recreational fishing are

allowed in the sanctuary, regulations have been established which prohibit various other activities such as sand and gravel mining. The sanctuary is a managed resource area equivalent to MPA Category VI (Hoyt, 2011; IUCN, 1994). The SBNMS encompasses only a small part of the Gulf of Maine sub population's summer range, and although some individuals are seen regularly there during the summer, none are thought to remain permanently within its boundaries.

1.2.3 Heterogeneity

Two known sources of heterogeneity for North Atlantic humpback whales are site fidelity and sighting probability. Many marine mammals, especially some baleen whales such as North Atlantic humpback whales, grey and southern right whales, are known to display a high degree of maternally directed fidelity to both feeding and breeding grounds (Clapham et al., 1993; Clapham and Mayo, 1987; Valenzuela et al., 2009). It is plausible that individual North Atlantic humpback whales vary in their propensity to use the marine sanctuary as opposed to other parts of the summer feeding range (Stevick et al., 2003). Heterogeneity in sighting probability is a well known phenomenon (Hammond, 1986, 1990) as the probability of sighting relies on individual behaviour at the beginning of a dive. The angle to which an individual's fluke shows on diving determines the probability of a successful photograph. The probability of being seen and recognized is therefore a combination of the true observation error (e.g. the randomness in viewing flukes) and also the real biological signal arising from individual heterogeneity in presence and absence in the SBNMS. With several sources of heterogeneity and many long individual capture histories (up to 27 years), investigation of this data set provides an excellent case-study for the application of the

methods of studying heterogeneity which are developed in this thesis.

The difficulty here is determining the underlying behaviour of the whales: observations are indications of presence in the SBNMS but whales may also be present and not observed. In addition, there is evidence of variation in individual annual arrival times, length of stay and usage of the marine park (Figure 1.2). Some whales are seen only a few times throughout the whole season; for others, a sequence of observations is followed by no observations for some time, followed by more observations - suggesting these whales are leaving and returning to the SBNMS. Population-level variation in use of the marine park is evident in both Figure 1.2 and Figure 1.3. Shared, population-level features are apparent, with an absence of dots evident for all whales in early 2001 (Figure 1.2). Note that effort is almost continuous. This thesis aims to uncover some of the underlying latent individual heterogeneity, but it is not realistic to expect that all aspects of the complex and annually-variable presence-absence cycle will be covered. The goal is to develop the statistical machinery to estimate individual heterogeneity, whilst incorporating some level of realism.

1.3 Multi-state mark-recapture models

One method to capture heterogeneity in probability of observation and site-fidelity is through the development of multi-state mark-recapture models. Multi-state mark-recapture models, first developed by Arnason (Arnason, 1972, 1973), extend traditional mark-recapture models by allowing animals to be in different 'states' (Lebreton et al., 2009). The 'movement' parameter, the probability of transitioning between states, was originally introduced to distinguish between emigration and mortality. The states, which may or

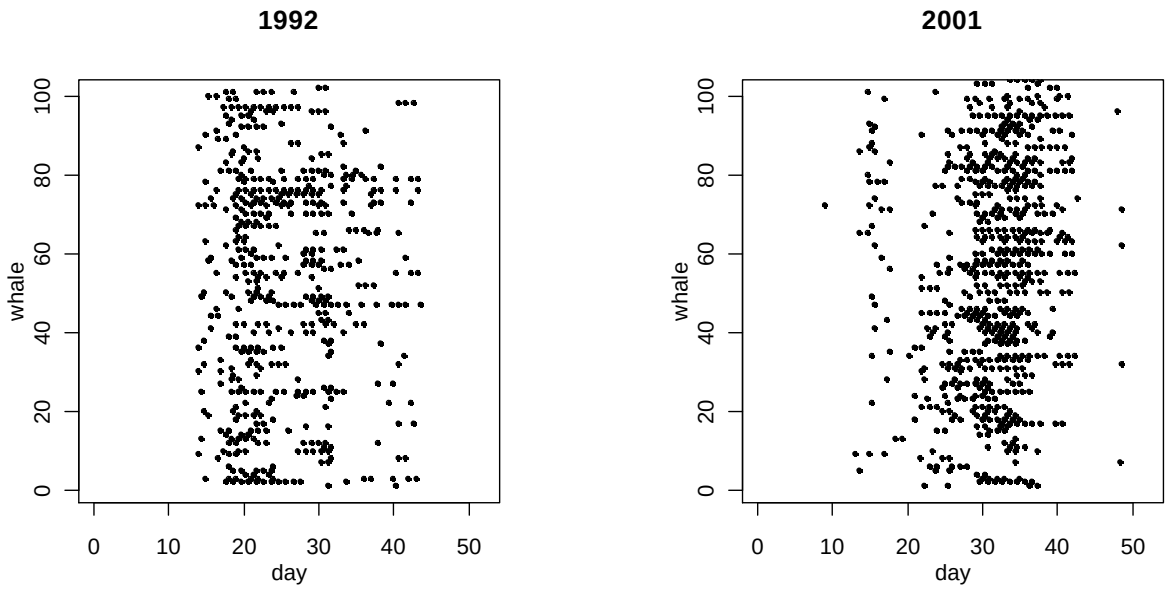

Figure 1.2: Daily record of observations for 100 individuals in years 1992 and 2001. Dots represent sightings.

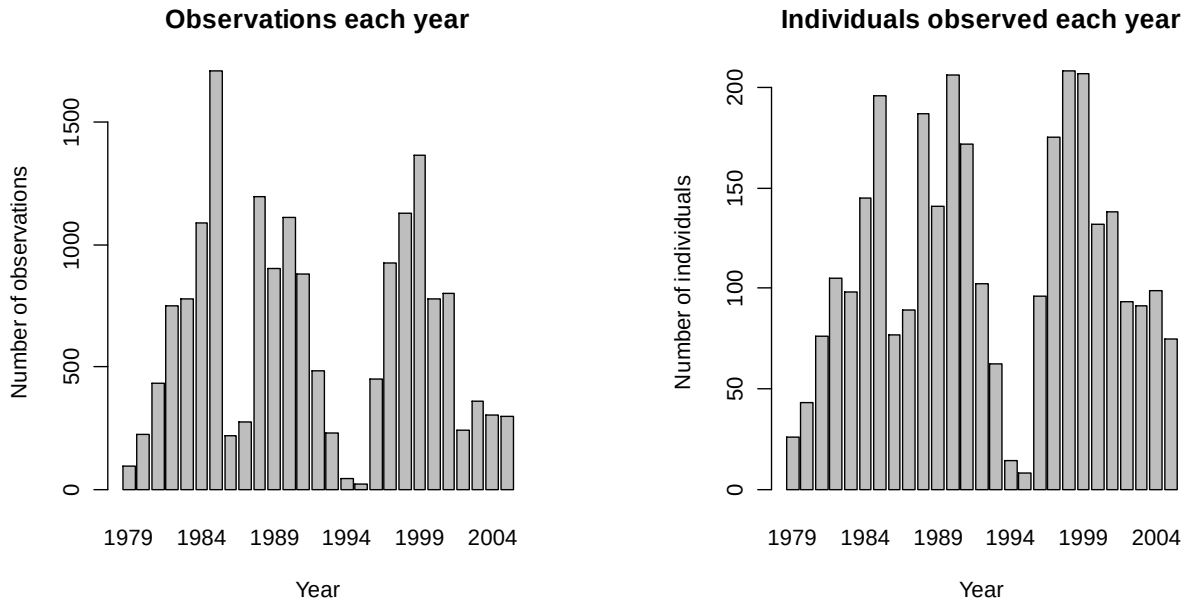

Figure 1.3: Total number of observations and number of individuals observed each year from 1979 through to 2005.

may not be directly observable, can include, but are not limited, to breeding, location, and behaviour. State can affect the probability of observation, and this can be built into the multi-state framework (Lebreton and Pradel, 2002).

Multi-event models (Pradel, 2005), extend the multi-state models through the use of a hidden Markov model framework, which incorporate a more realistic assessment of natural events as transitions between states can be treated as a Markov process that is not directly observable (Conn and Cooch, 2009; Zucchini et al., 2008). In hidden Markov models, two time series - the observation and process components - run in parallel (Gimenez et al., 2012). The observation process (e.g. seen/not seen) does not usually reveal the current underlying state directly, but does provide indirect information on the probable state (e.g. present in the marine park, not present, or dead). Modeling both the process and observation component enables the separation of the real signal from the observation error (Patterson et al., 2008).

In this thesis I extend a three state hidden Markov model to allow for individual variation on both the observation and process components. Complexity arises partly through the application of a hidden Markov model with population-level fixed effects, but mainly through the incorporation and estimation of individual heterogeneity.

1.4 Techniques for estimating individual heterogeneity

In situations where individual heterogeneity is thought to be important, it is not always immediately clear how to model it. Various techniques exist for addressing latent individual heterogeneity and generally involve one of

the following assumptions regarding the underlying nature of the individual heterogeneity:

1. No individual variability within the population.
2. Individual fixed parameter value which can't usefully be predicted from knowing the values for other animals.
3. Individuals may fall into one of a number of fundamentally distinct groups. Within each discrete group, individuals are assumed to behave in the same way (i.e. to have the same parameter value).
4. Individuals have a unique parameter value which is sampled from a specific parametric distribution, e.g. $N(\mu, \sigma^2)$ a Normal distribution with mean μ and standard deviation σ^2.
5. A model in which individuals cluster into a potentially infinite number of groups.

If all animals have very short capture histories, there may be little gain in modeling heterogeneity at all. This is because there will likely be not enough data to estimate it, and it may only be practical to acknowledge qualitatively the possibility that the unaccounted heterogeneity has influenced the other results. In this case, method 1 would be the most appropriate. On the other hand, if all animals have sufficiently long capture histories, then heterogeneity could easily be handled using fixed effects (method 2).

Within mark-recapture research, Pledger (2000) and Norris and Pollock (1996) developed finite mixture models (method 3) which use a discrete-valued prior supported on a prespecified number of points. The location of these points, i.e. the allowable parameter values, and the proportion of the population associated with each point constitute the parameters to be

estimated. Despite the computational advantage over other methods (e.g. method 4), this approach relies on the assumption of a prespecified number of groups and can result in model selection issues through both the choice of the number of mixture components and the parameters characterizing them (Cubaynes et al., 2012).

Individual-level random effects using continuous distributions (method 4) are a more natural candidate for modeling heterogeneity, where the variance reflects between individual variability. However, despite being the focus of much research (Maunder et al., 2008; Lebreton et al., 2009; Barry et al., 2003; Royle and Dorazio, 2008; Huggins and Yip, 2001; Burnham and Overton, 1978; Gimenez and Choquet, 2010), methods for the inclusion of multiple levels of continuous individual-level random effects into mark-recapture models are still not commonly used. This is in part due to the complex calculation required to deal with these random effects.

Random effects can be used to model a proportion of variance arising from persistent unobserved individual heterogeneity. The problem is that in order to estimate this variance within a population, it is necessary to integrate across all possible values of the individual-level random effects which is not straight forward. Taking $y_i = \theta + u_i$, for data y_i, fixed population-level effect θ, and individual-level random effect u_i, then the likelihood function for θ is $L(\theta) \propto g(y, \theta, u)$, where $g(y, \theta, u)$ is the joint density of y, θ, and u. To make inferences of θ given y we must first integrate out u, which is generally computationally difficult and cannot be done analytically, so an approximate technique is required.

Several methods are available for parameter estimation in a mixed model, including penalized quasi-likelihood (Breslow and Clayton, 1993), Laplace approximation (Raudenbush et al., 2000; Tierney and Kadane, 1986), Gauss-

Hermite quadrature (Gimenez and Choquet, 2010; Liu and Pierce, 1994) and Markov Chain Monte Carlo (MCMC) (Gilks et al., 1996). The methods I develop in this thesis explore the use of both Laplace approximation and MCMC.

In this thesis I explore the incorporation of multiple measures of individual heterogeneity using methods 4 and 5. In parametric models (method 4), the functional form of the prior distributions are always assumed to be known. In contrast, non-parametric model (method 5), such as the Dirichlet process, can allow individuals to belong to one of a potentially infinite number of groups. The natural subgroupings often seen in mark-recapture studies and the complexity of real mark-recapture data means that parametric models can be insufficient. The Dirichlet process provides a method to allow identification of a number of subgroups without prespecifying the number of groups *a priori*. A parametric model is developed in chapter 4, and the non-parametric Dirichlet process is described, and implemented, in chapter 5.

1.5 Thesis structure

The primary focus of this thesis is the development of methods which improve the effectiveness of capturing individual variability in mark-recapture data. Although applied to a population of North Atlantic humpback whales, the methods extend to other marine species and also to the terrestrial world. Through the quantification of individual heterogeneity, and the application to MPAs, these methods have the potential to influence management decisions.

In chapter 2 I demonstrate, using a simulation study not unlike the

case-study, the population consequences of individual heterogeneity in the context of a MPA. This application illustrates the potential importance of individual variation in a conservation context.

In chapters 3, 4, and 5 of this thesis I address the problem of incorporating individual heterogeneity into mark-recapture models. I develop several techniques and approaches to the same problem (Figure 1.4). Using a hidden Markov model framework, I assume that the individual observation histories are indicative of latent behaviour of movement in and out of a marine reserve (e.g. SBNMS).

In chapter 3 I incorporate individual variability using an implementation in the software Automatic Differentiation Model Builder (ADMB) which is generalisable and easily used. I use simulations to investigate the costs and trade-offs associated with fitting too complex or too simple a model by testing the effect of model misspecification (i.e. mistakenly assuming no latent heterogeneity, or assuming latent heterogeneity when none is in fact present). In application to North Atlantic humpback whale data, I incorporate three levels of individual heterogeneity, plus annual population-level variation in probability of using the marine park. I explore one derived demographic parameter of the method which quantifies the time spent in the SBNMS.

In chapter 4 I combine work by Zucchini et al. (2008) and Scott (2002) to develop a Beta-Binomial MCMC sampler for the hidden Markov model. An Independent Metropolis-Hastings sampler is developed to allow efficient updating of the hyper-parameters which cannot be updated using Gibbs sampling.

In chapter 5, I develop the method further using a non-parametric Dirich-

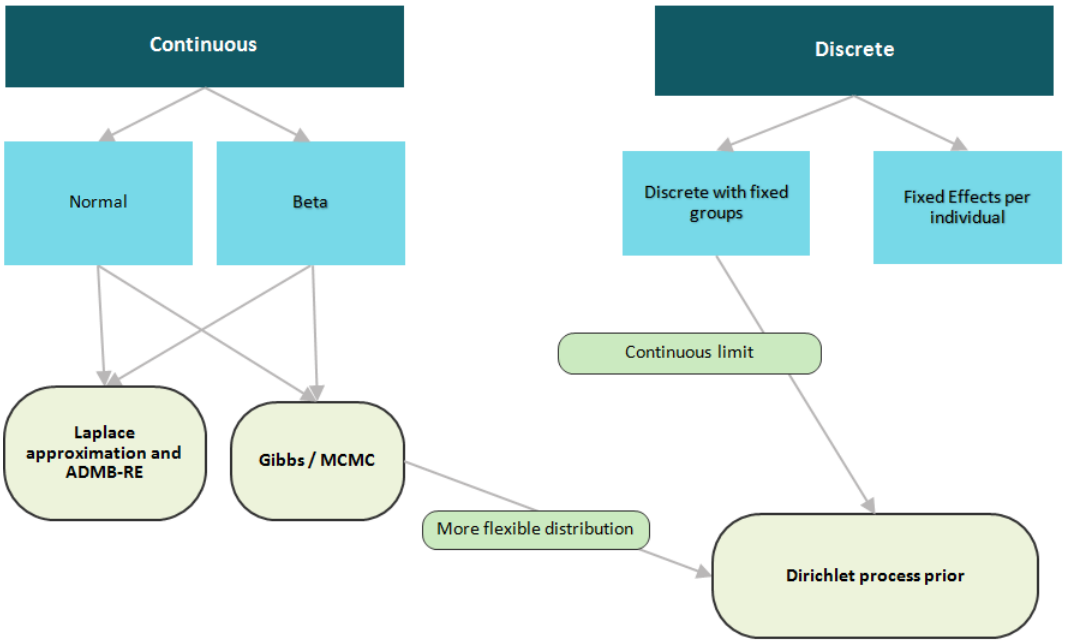

Figure 1.4: Pathways to modeling individual heterogeneity.

let process prior. This approach extends the Beta-Binomial model to a more flexible distribution, avoiding assumptions of unimodal behaviour, and extends the discrete groups approach (Pledger, 2000) to the continuous limit. I develop a sampler based on work by Neal (2000), which allows for multimodality in latent individual variation in both the observation and process components. I also combine work by Murugiah and Sweeting (2012) and Escobar and West (1995) to update the precision parameter of the Dirichlet process.

Finally, chapter 6 summarises conclusions and explores possibilities for future work.

Chapter 2

The importance of individual heterogeneity in the effectiveness of marine protected areas.

Summary

The focus of this thesis is the development of methods to estimate individual heterogeneity in mark-recapture analysis. This chapter illustrates that accounting for individual heterogeneity is not merely a statistical concern, but that individual heterogeneity has potential population consequences in the context of marine spatial management. A simple population dynamics model is used to explore the effectiveness of a marine protected area (MPA) given differing levels of population coverage, individual heterogeneity in site fidelity and differential survival inside and outside the MPA. There is a huge amount of literature on MPAs as a management tool, but to the best of my

knowledge there has been little consideration of individual heterogeneity in estimating the effectiveness of MPAs.

2.1 Introduction

Declines in marine mammal populations have been attributed to a variety of factors: entanglement in fishing gear (Hoyt, 2011; Johnson et al., 2005; D'Agrosa et al., 2000), over fishing (Read et al., 2006), pollution (Wilgart, 2007), and ship strikes (Laist et al., 2001; Knowlton and Kraus, 2001). As these threats are often concentrated in space, MPAs have been advocated as an effective management strategy for mitigation. Determining the effectiveness of a MPA's spatial configuration (e.g. location, extent) is challenging. An often overlooked factor is how individual heterogeneity in spatial use can mediate the effectiveness of a MPA.

There are presently over 500 MPAs primarily for marine mammals. Marsh (2000) discussed the use of a MPA for Dugongs (*Dugong dugon*) and the implications of the threat of bycatch outside the protected area. Residency appears to be relative to the quality of the habitat: if the quality of the habitat within the marine protected area degrades, then the animals are more likely to forage further afield, resulting in higher proportion of the population being at risk to bycatch (Marsh, 2000).

MPAs rely on the notion that threat to an exposure is reduced inside the protected area. The design of MPAs will vary depending on the target species. For some species (e.g. abalone, rock lobster), the concept of home range will likely be the defining feature; previous studies have established the link between home-range and MPA size in determining effectiveness of a MPA (Gell and Roberts, 2003; Moffitt et al., 2009). MPA design for mi-

gratory species (e.g. cetaceans) will focus, for example, on the protection of critical habitat (Williams et al., 2009), or through the mitigation of bycatch (Gerrodette and Rojas-Bracho, 2011). For mobile species, an important, but much overlooked consideration in determining the effectiveness of a MPA is the inherent variation in spatial use by individuals within a population (as discussed by Kerwath et al. 2008; Kaplan et al. 2010; Moffitt et al. 2009, 2011).

This chapter focuses on the implications of individual heterogeneity in site fidelity to a MPA; estimation of this individual heterogeneity is covered in chapters 3, 4 and 5 of this thesis. These methods rely on individual-level data obtained using mark-recapture techniques, now commonly used within marine ecology (Beausoleil et al., 2004). Despite the frequency with which individual-level data is collected, and that individual heterogeneity in populations is known, I was unable to find examples in the literature of consideration of the population consequences of individual heterogeneity in the context of MPAs. To the best of my knowledge there has been only limited reference to individual variation for existing MPAs (e.g. Kerwath et al., 2008). This may be due to a lack of methods. If methods were available, two cases in particular which have appropriate data and where applications would be interesting, are described in papers by Williams et al. (2009) and Gormley et al. (2012). Williams et al. (2009) looked at social aggregations and protected areas in killer whale conservation, and Gormley et al. (2012) explored the effectiveness of the Banks Peninsula Marine Mammal Sanctuary for Hectors dolphins.

The goal of this chapter is to provide an illustration of the population consequences of individual heterogeneity in spatial use in the context of a MPA. Individual heterogeneity in spatial use is explored, using as a proxy,

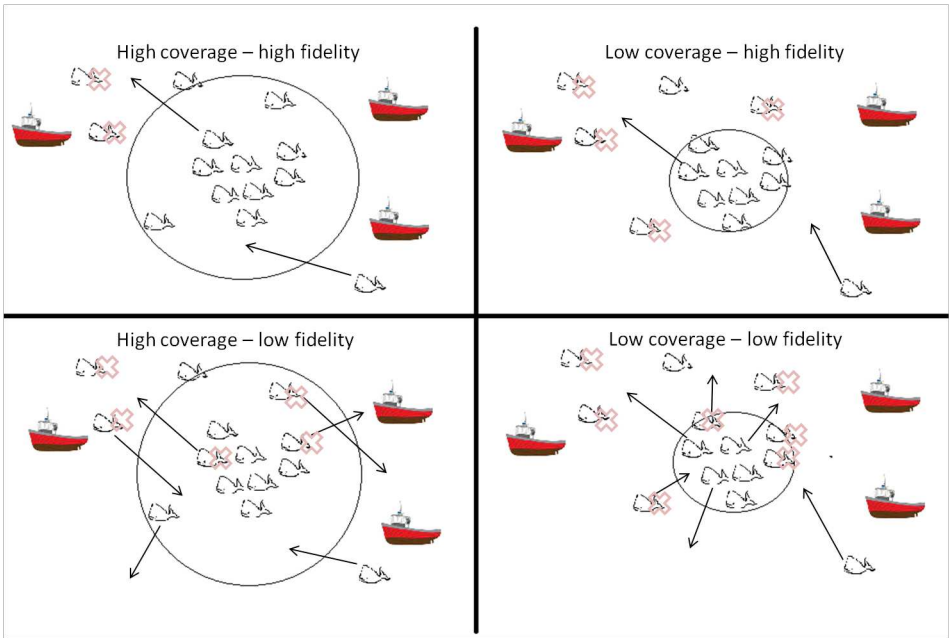

Figure 2.1: The relationship between population-level exposure to threats (coverage) and individual-level exposure (fidelity) given differential mortality outside a MPA. Arrows indicate the amount of movement in the population. Low fidelity to an area and lots of movement will result in more movement in and out of the protected area and thus higher proportion of time exposed to the threat and greater chance of death (indicated by light red crosses).

the proportion of time spent in the reserve. We explore the effectiveness of a MPA given differing levels of two factors we term coverage and fidelity where:

- coverage is mean population-level proportion of time spent inside the protected area and

- fidelity is the proportion of time an individual spends inside the protected area.

Simulations are used in this chapter to explore the population consequences of individual heterogeneity in proportion of time spent in the reserve. If there is a lot of movement within a population, and large differential survival (between inside and outside a protected area), then all animals will be somewhat exposed; if there is less movement, there will be less exposure, and thus less mortality from the threat (see Figure 2.1). The impact of heritable site fidelity is also explored, as maternally directed site fidelity can affect the time taken to recolonize certain areas for extirpated segments of populations (Turgeon et al., 2012). Many marine mammals, especially baleen whales such as North Atlantic humpback whales and southern right whales, are known to display a high degree of maternally directed fidelity to both feeding and breeding grounds (Clapham et al., 1993; Clapham and Mayo, 1987; Valenzuela et al., 2009).

The basic population dynamics model and methods used to incorporate individual heterogeneity and heritability are described in section 2.2. The results of the population projections are presented in section 2.3, and following this in section 2.4 is a discussion of results and implications.

2.2 Methods

An individual based discrete-time population dynamics model was used to simulate population-level consequences of differential survival and individual heterogeneity in proportion of time spent inside and outside a marine reserve (which we refer to as *fidelity*, symbolised by h, or 'hereness'). We consider two levels (high/low) for each of the following three factors: population-level coverage; heterogeneity in individual-level fidelity to the MPA; and survival outside the protected area. The simulated populations were assumed to consist of only females, and all individuals were able to breed from age 1. Note that the model was kept simple in order to focus on the interaction between coverage, individual heterogeneity in fidelity, and differential survival.

Time steps were assumed yearly and each year an individual's location was determined by their fidelity, with probability h_i of being inside the MPA. Survival was then determined by location: an individual survived the year with probability S_I if inside the reserve and S_O if outside. Given that protection is afforded by a reserve, it was assumed that $S_I > S_O$. The population, N_t, in year t, is given by the number alive at the end of the year (those remaining alive plus the number of births).

Breeding was assumed to be density dependent: individuals breed with probability $c_1 \times (1 - \frac{N_t}{K}) + c_2$, where K is carrying capacity ($K = 100$ for all projections), defined jointly across inside and outside. The value c_2 was chosen to ensure the population was in equilibrium when $S_O = S_I$, so that $c_2 = 1 - S_I$ i.e. there was no extra mortality outside the MPA. The constant c_1 was calculated as: $c_1 = 0.1 - c_2$. This means the per capita breeding rate varied between 0.1 at $N_t = 0$ and 0.05 at $N_t = K$, when $S_I = 0.95$.

Two scenarios for MPA coverage were considered (Figure 2.2): high cov-

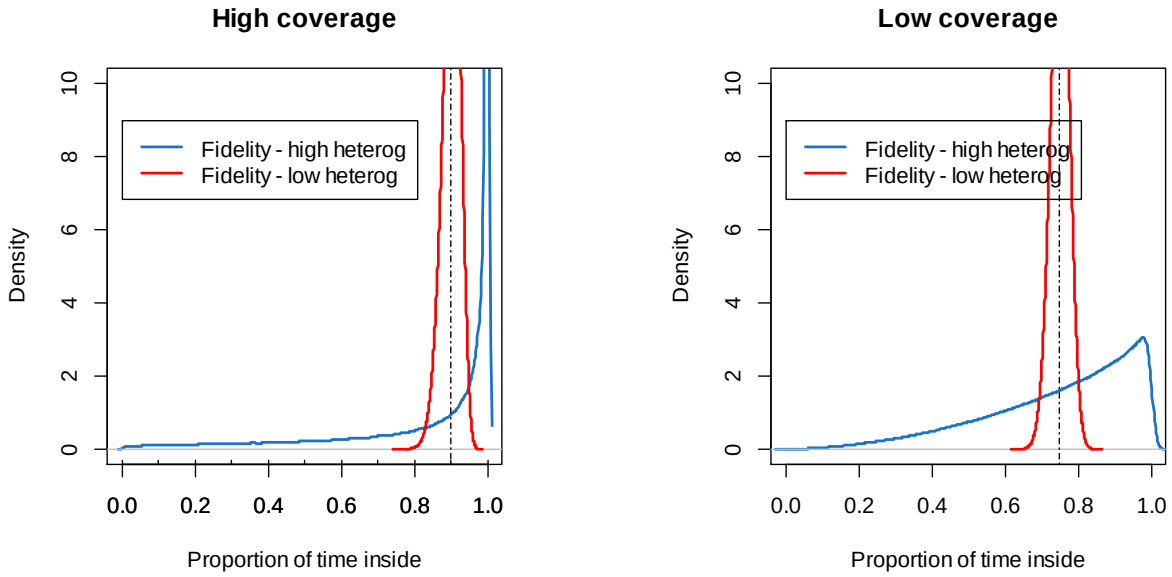

Figure 2.2: (a) high population-level coverage (mean population coverage of 0.90, indicated by the thin black dotted vertical line) with two levels of heterogeneity in fidelity; (b) low population-level coverage (mean population coverage of 0.75, indicated by the thin black dotted vertical line) with two levels of heterogeneity in fidelity.

erage with a population-level mean fidelity of 0.90; and low coverage with a population-level mean fidelity of 0.75. Likewise, two levels of individual heterogeneity in fidelity were explored using two Beta distributions with the same mean, but different variance: high heterogeneity in fidelity using a variance of 0.04, and low heterogeneity in fidelity using a variance of 0.000625 (see Figure 2.2). For the purpose of simulation, an individual's fidelity was sampled from a Beta distribution, $h_i \sim Beta(a, b)$ where parameters a and b were chosen to give the desired mean and variance. One extension of this model which was explored was the population consequences of heritable fidelity where an individual inherits their mother's fidelity.

Two scenarios for survival outside the MPA ($S_O = 0.90, 0.80$) were used to explore the importance of accounting for individual heterogeneity when quantifying MPA effectiveness. In each scenario $S_I = 0.95$. Simulations were initialized with population of $N_0 = K$ (100) and run for 1000 years to enable population trajectories to reach equilibrium. Each scenario was run 500 times.

2.3 Results

High Coverage: As expected, high population-level coverage resulted in stable population trajectories for all four scenarios (high/low fidelity and high/low S_0) (see Table 2.1). After initial decline in population size (largest decline for low fidelity, low S_0), all four scenarios appeared to reach equilibrium at around 200 years. High heterogeneity in fidelity and high S_0 resulted in the highest population size. High S_0 resulted in larger population size at equilibrium than low S_0. The numbers inside the protected area displayed similar pattern to total population for all four scenarios. However, the pat-

a)

Total Population			**Total Population**		
High coverage	High S_0	Low S_0	**Low coverage**	High S_0	Low S_0
High Var(h_i)	0.98	0.79	High Var(h_i)	0.72	0.12
Low Var(h_i)	0.86	0.62	Low Var(h_i)	0.69	0.002

b)

High coverage	High S_0	Low S_0	**Low coverage**	High S_0	Low S_0
High Var(h_i)	0.95	0.81	High Var(h_i)	0.72	0.35
Low Var(h_i)	0.83	0.60	Low Var(h_i)	0.65	0.19

c)

High coverage	High S_0	Low S_0	**Low coverage**	High S_0	Low S_0
High Var(h_i)	0.96	0.39	High Var(h_i)	0.61	0.24
Low Var(h_i)	0.87	0.59	Low Var(h_i)	0.67	0.18

Table 2.1: Tables indicate the mean proportion of the population alive after 1000 years (the mean of 500 projections) for: (a) total; (b) inside; and (c) outside.

tern was different for the number outside the protected area: at low S_0, low heterogeneity resulted in larger population size outside than high heterogeneity (Figure 2.3 and Table 2.1). (Figures for total population size and population inside are not presented due to their similarity in trajectories.)

Low Coverage: For low coverage, there was little effect of heterogeneity at high S_0. However, for low S_0 the population decline and time to extinction (based on visual examination) was quicker with low heterogeneity (Figure 2.4). This was not evident with high heterogeneity due to the replenishment of the outside population from the high rate of movement across the boundaries (between inside/outside) of the MPA. For the numbers inside, the overall pattern was similar to the trajectories of total population.

The pattern was different for the number outside the protected area: at high S_0, low heterogeneity resulted in larger population size outside than high heterogeneity (Figure 2.3 and Table 2.1).

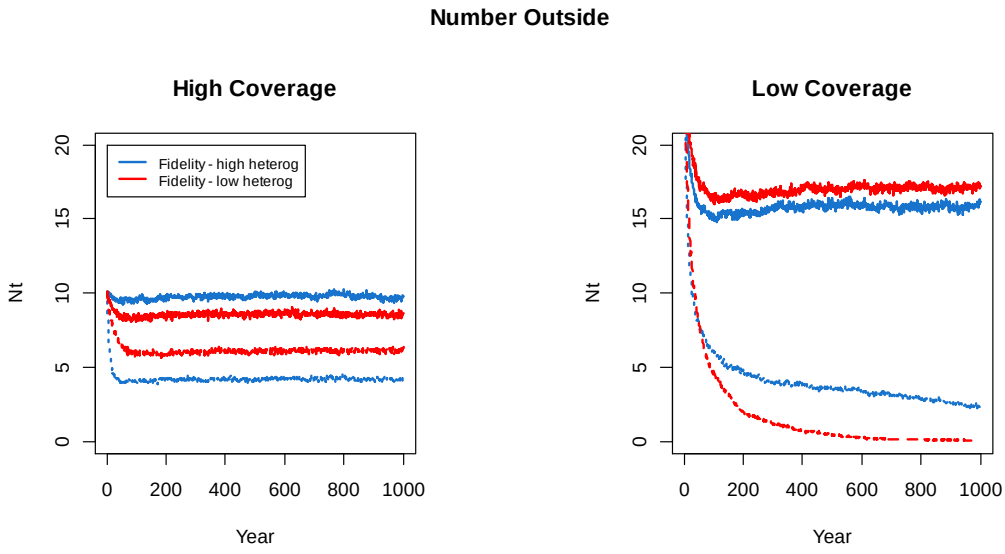

Figure 2.3: The number outside for both High and Low coverage for two levels of heterogeneity in fidelity and S_0. Solid lines show conditions for $S_O = 0.90$ and dashed $S_O = 0.80$.

Heritable Fidelity: Heritable fidelity resulted in clearly different trajectories compared to non-heritable fidelity. For several scenarios, an initial decline was followed by an increase in population size, resulting in longer time to reach equilibrium.

For both high and low coverage (and high/low S_0), projections for high heterogeneity in fidelity indicate initial declines in population size followed by an increase in the population to equilibrium (see Figure 2.5 for low coverage). This is due to the selection, through heritable fidelity, of individuals with high fidelity to the protected area. Equilibrium for heritable fidelity was not reached until around 500 years; in comparison to around 200 years for non-heritable fidelity.

For low coverage, low heterogeneity, and high S_0, there was an initial population decline followed by equilibrium, but no increase in population size. Despite heritable fidelity, the combination of low coverage, low S_0 and

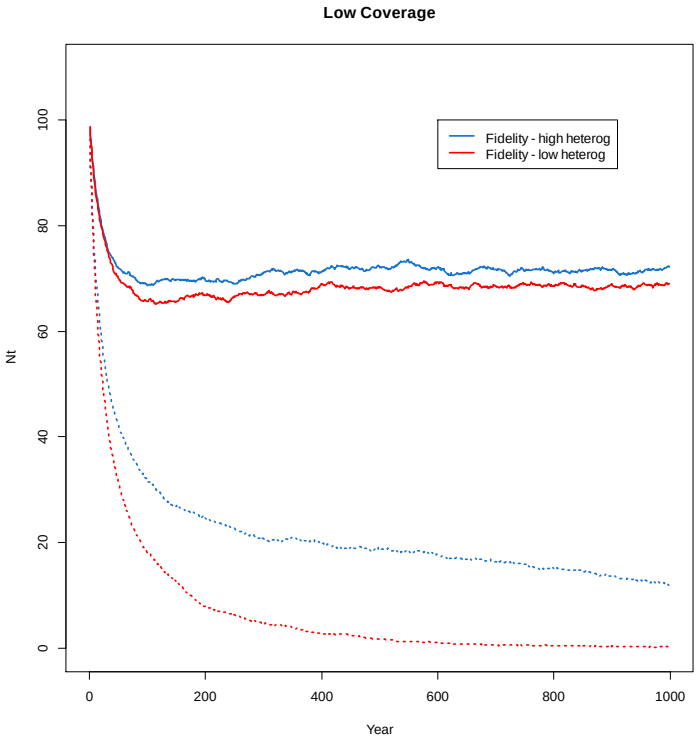

Figure 2.4: Mean Population size for 500 projections for 1000 years for low coverage with two levels of heterogeneity in fidelity - high and low. Solid lines show conditions for $S_O = 0.90$ and dashed lines $S_O = 0.80$.

a)

Total Population			**Total Population**		
High coverage	High S_0	Low S_0	**Low coverage**	High S_0	Low S_0
High Var(h_i)	0.98	0.98	High Var(h_i)	0.94	0.90
Low Var(h_i)	0.89	0.72	Low Var(h_i)	0.70	0.01

b)

Inside			**Inside**		
High coverage	High S_0	Low S_0	**Low coverage**	High S_0	Low S_0
High Var(h_i)	0.98	1.08	High Var(h_i)	0.90	0.74
Low Var(h_i)	1.22	1.18	Low Var(h_i)	0.71	0.01

c)

Outside			**Outside**		
High coverage	High S_0	Low S_0	**Low coverage**	High S_0	Low S_0
High Var(h_i)	0.98	0.01	High Var(h_i)	0.74	0.51
Low Var(h_i)	0.13	0.07	Low Var(h_i)	0.66	0.01

Table 2.2: Tables indicate results for heritable fidelity. The mean proportion of the population alive after 1000 (the mean of 500 projections) years for: (a) total; (b) inside; and (c) outside.

low heterogeneity resulted in continuous population decline (Figure 2.5). The presence of heritable fidelity resulted in much higher proportion of the population inside the protected area at equilibrium; this was due to the selection of individuals favouring the protected area (Table 2.2).

Depending on the combination of high/low S_0 and high/low coverage, then outside the protected area was either better or worse off depending on variance of heterogeneity: low heterogeneity resulted in larger population size outside the protected area for high coverage/low S_0; high heterogeneity resulted in larger population size outside for low coverage/high S_0 (Table 2.2).

2.4 Discussion

The modeling shown here indicates that individual heterogeneity in spatial use and site fidelity could have important implications under certain con-

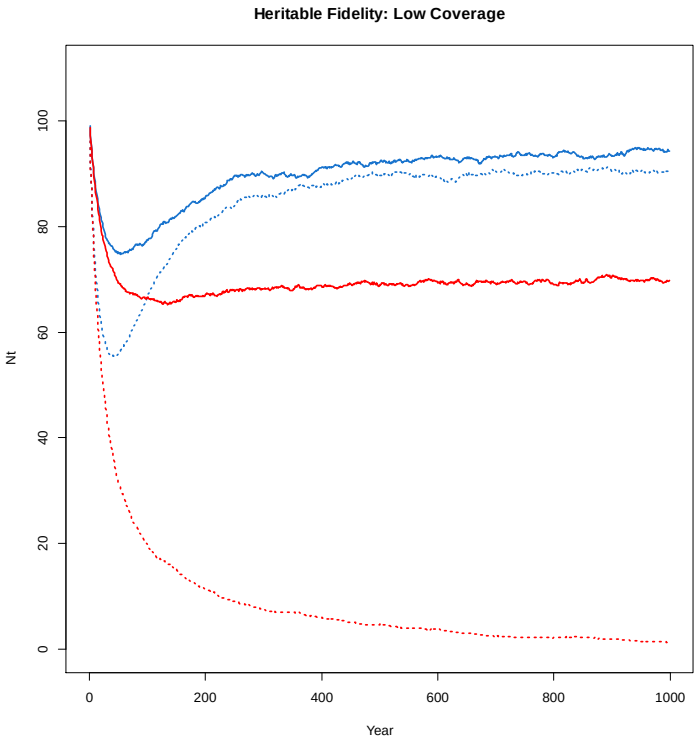

Figure 2.5: Mean Population size for 500 projections for 1000 years for heritable fidelity for low coverage with two levels of heterogeneity in fidelity - high and low. Solid lines show conditions for $S_O = 0.90$ and dashed lines $S_O = 0.80$.

ditions for the dynamics of populations managed using MPAs. In several scenarios, high individual heterogeneity resulted in larger population size and positive population trajectories, compared to decline and eventual extinction with low individual heterogeneity in site fidelity. The patterns for population size outside the MPA were interesting and differed depending on the combination of differential survival, coverage and heterogeneity in fidelity. In practice, MPAs will often cover a smaller proportion of area and animals than in these simulations, but these projections illustrate the importance of accounting for what happens to the population outside the boundaries of a protected area.

The results highlight that although a population might appear to be healthy, a portion of the population may be killed off. For most projections, high heterogeneity in fidelity resulted in larger population size at equilibrium. Two exceptions to this were when low heterogeneity in fidelity resulted in larger population size outside the protected area. The conditions when this occurred were with: high coverage and low survival outside; and low coverage and high survival outside.

The selection of individuals favouring the protected area, due to heritable fidelity, resulted in a larger proportion of the population inside the reserve at 1000 years in comparison to projections with non-heritable fidelity. Interestingly, for both high and low coverage, positive population growth (after the initial decline) was only observed for high heterogeneity in heritable fidelity.

For cetaceans, space use varies from extensive ocean wide migrations (e.g. fin whale *Balaenoptera physalus*, humpback whales *Megaptera novaengliae*), through to ranges over a small geographic area (e.g. Vaquita *Phocoena sinus*) (Hoelzel, 1998). The effectiveness of MPAs for wide-ranging or non home-range animals has been questioned (Hoyt, 2011). However, if a high

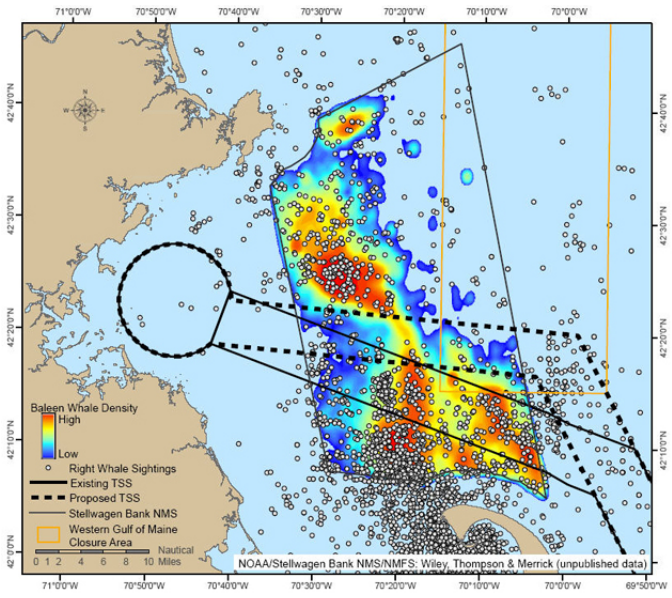

Figure 2.6: . The Stellwagen Bank National Marine Sanctuary and surrounding waters showing the distribution and relative density of all baleen whales in the sanctuary, the location of right whale sightings by "o" and the current and proposed Traffic Separation Schemes through the sanctuary. New shipping lanes indicated by the dashed lines. Data consist of over 350,000 sightings over a 24 year period. (Source: SBNMS(2006)).

proportion of the population is exposed to a spatially-concentrated threat at some stage, for instance in a narrow migration corridor, then even a small MPA may be effective if well chosen (Kaplan et al., 2010; Kerwath et al., 2008). For many cetaceans, the design of MPAs focuses on the protection of critical habitat such as feeding, breeding, and calving grounds (Hoyt, 2011). One example of the effectiveness of changes made to spatial threats within a feeding ground is the rerouting of shipping lanes crossing the Stellwagen Bank National Marine Sanctuary. Even the small changes (increasing traffic time by only 10-22 minutes), resulted in an expected reduction in risk of collision in the SBNMS by 58% for the critically endangered Right whale, and 81% for all other baleen whales (e.g. North Atlantic humpback whales) (Stellwagen Bank National Marine Sanctuary, 2006).

Following the results of the population simulations in this chapter, the remaining chapters in this thesis focus on the development of methods to better detect and estimate the degree of individual heterogeneity: a Markov model describing the transition between In and Out (and death) is estimated, and three statistical methods for modeling (and detecting) individual heterogeneity are developed and contrasted. Individual heterogeneity in these parameters are of interest in their own right, but also of importance in order to avoid bias in estimates of survival, abundance and movement.

This chapter has been removed for copyright or proprietary reasons.

CHAPTER 3

Incorporating individual variability into mark-recapture models

Published in

http://onlinelibrary.wiley.com/doi/10.1111/j.2041-210X.2012.00243.x/abstract

Ford, Jessica H, Bravington, Mark V, Robbins, Jooke. (2012). Incorporating individual variability into mark-recapture models. *Methods in Ecology and Evolution 3 (6) 1047-1054.*

Chapter 4

A Markov chain Monte Carlo sampler for a hierarchical hidden Markov mark-recapture model.

4.1 Introduction

The focus of this chapter is the development of a Markov chain Monte Carlo (MCMC) sampler for a hierarchical Bayesian hidden Markov model, applied to mark-recapture data, which allows for individual heterogeneity in both the observation and process components of the model. It is intended as an alternative approach to the Laplace approximation in chapter 3 and to provide sections of the MCMC sampling framework required for the more complex non-parametric model developed in chapter 5.

The structure of the model in this chapter is the same as used in the

previous chapter: a hidden Markov model which allows for individual variability on the probability of observation and the probability of remaining in either of the two states Here and Away. Individual-level random effects are assumed to follow a Beta prior distribution; given the data, this choice of prior means that the individual-level effects have posterior distributions that are also Beta, and can be updated simply via Gibbs sampling. The hyper-parameters of the Beta distribution are updated using a changing Independent Metropolis-Hastings sampler; and the population-level fixed effects are updated using an unchanging Independent Metropolis-Hastings sampler derived from a preliminary Laplace approximation.

Section 4.2 briefly describes some of the MCMC theory used in this chapter. The basic two-state model used for simulation testing is outlined in section 4.3, and the computational steps involved in the MCMC routine in section 4.4. In section 4.5 I test for asymptotic convergence of individual parameter values and population-level hyper-parameters using the two-state model. The full three-state model (including death and using the same structure as chapter 3), is applied to North Atlantic Humpback whales in section 4.6. Parts of the methods are reiterated from chapter 3 for clarity.

4.2 MCMC

Markov chain Monte Carlo provides a solution to the calculation of marginal distributions which involves complex integrals. The main appeal of this approach is that it provides a powerful and flexible framework for analysing complex stochastic processes that would be intractable by other methods.

Thomas Bayes (1763) first proposed Bayes Theorem which underpins all Bayesian inference. Bayes theorem defines the joint posterior distribution

for the data, x, and all other parameters in the model, θ, as

$$p(\theta|x) = \frac{p(x|\theta)p(\theta)}{p(x)}$$

where $p(\theta)$ is the *prior*, $p(\theta|x)$ the *posterior* and $p(x|\theta)$ the *likelihood*. For most models it is simple to determine the posterior density $p(\theta|x) \propto p(x|\theta)p(\theta)$ to within a multiplicative constant. However, the calculation of the normalisation constant $\int p(x|\theta)p(\theta)d\theta$ is computationally demanding, and in some cases the log-likelihood itself is not easily computable, as it involves integration over many hidden variables (e.g. the state chains in a hidden Markov model). The appeal of MCMC lies in the ability to generate samples from a posterior even when the normalising constant is intractable, which is often the case except for all but the most simple models. MCMC works through simulation by sampling draws from a Markov Chain constructed so that $p(\theta|x)$ is is its equilibrium distribution. Successive draws are not independent, so a large number of draws may be required to overcome the autocorrelation. In order to draw these samples we require both a method for proposing what the next draw might be, and a method for deciding whether to accept or reject the draw; if a proposed draw is rejected, the chain stays at its previous state.

The Metropolis-Hastings Algorithm (Metropolis et al., 1953; Hastings, 1970) is the most generic form of MCMC: given a target distribution p, the Metropolis-Hastings algorithm constructs a Markov chain which has p as its stationary distribution. To generate a sequence of draws, $\theta_1, \theta_2, \theta_3, ...$, the candidate values are generated from the proposal distribution $q(\theta^\circ; \theta^*)$ where θ^* represents the current parameter value, and θ° the proposed value. Updating for the next iteration, θ, involves accepting or rejecting the pro-

posed candidate value according to

$$\theta = \begin{cases} \theta^{\circ} \text{ with probability } \min\left\{1, \frac{p(\theta^{\circ})q(\theta^{*};\theta^{\circ})}{p(\theta^{*})q(\theta^{\circ};\theta^{*})}\right\}, \\ \theta^{*} \text{ otherwise.} \end{cases}$$

Standard random-walk Metropolis-Hastings algorithms (Metropolis et al., 1953; Hastings, 1970) begin with an arbitrary value, and follow by using the proposed jumping distribution which is usually centered on the current value and will therefore move as the chain evolves. Candidate values are generated from this proposal distribution which generally takes the form of a density function. The speed of convergence of the MCMC (i.e. how many samples are required to get a reliable answer) is determined by the choice of proposal distribution. Often, the proposal distribution is tuned to get close to the desired acceptance rate, which on theoretical grounds has been suggested to be about 0.45 for Metropolis-Hastings algorithms. Another version of Metropolis-Hastings is known as Independent Metropolis-Hastings sampler. Here, the probability of a jump to the candidate value is independent of the current value (Keith et al., 2008). The candidate value should be drawn from a proposal distribution that is close to the target distribution. The reliability and efficiency of the Independent Metropolis-Hastings sampler relies on the proposal distribution: if it is well chosen then the Independent Metropolis-Hastings sampler is efficient with high acceptance rate (close to 1, in contrast to standard Metropolis-Hastings).

The Gibbs sampler, a special case of Metropolis-Hastings, generates samples from the posterior distribution by drawing from the full conditional distribution: the distribution of each parameter conditional on the data and all other parameters (Tierney, 1994; Gelfand and Smith, 1990; Gelfand, 2000).

4.3 Model

As usual with hidden Markov or multi-state models, the overall model is split into a process part and an observation part. For the process model, we assume that at time t an animal i can be in either of two states S_{it}: Here and Away, or H/A for short.

Changes in the state over time are governed by a Markov process with transition matrix γ, so (omitting dependence on i for now) for any two states s and s^* we have

$$\mathbb{P}\left[S_{t+1} = s^*\right] \quad = \sum_s \quad \gamma^{ss*} \mathbb{P}\left[S_t = s\right]$$

The four elements of γ can be written in terms of just two parameters γ^{HH} and γ^{AA} (respectively the probabilities of staying Here and staying Away), as follows:

$$\gamma \quad = \quad \begin{pmatrix} \gamma^{HH} & \left(1-\gamma^{HH}\right) \\ \left(1-\gamma^{AA}\right) & \gamma^{AA} \end{pmatrix}$$

For the observation model, there are "capture attempts" (photo-ID expeditions) at each t, in which an animal may be seen if and only if it is Here. Our data for animal i are thus a time series $X_{i,t_{1i}:T}$ of 0 s (not seen) and 1 s (seen) where t_{1i} denotes the first observation of the animal (see below) and T the most recent expedition. If $X_{it} = 1$ then we know $S_{it} =$ H ,

but if $X_{it} = 0$ the state cannot be determined for certain. Formally, the probability of observation given state is expressed in terms of a parameter π by

$$\begin{aligned} \mathbb{P}\left[X_{it} = 1 | S_{it} = s\right] &= \begin{cases} \pi_{it} & s = \mathrm{H} \\ 0 & s = \mathrm{A} \end{cases} \\ \mathbb{P}\left[X_{it} = 0 | s\right] &= 1 - \mathbb{P}\left[X_{it} = 1 | s\right] \end{aligned}$$

Although there is potentially some information in the "pre-history" of an animal before it is first seen, in practice there is some additional complexity involved in considering it. In order to consider the early life history of an animal, during which time they may or may not be seen, one needs to introduce a random birth date and extra parameters to account for life history and state. An animal may not be seen simply because it has not yet recruited to the population, and our focus here is not on the recruitment process. We therefore start each animal's series at its first sighting of the given year, and condition on $S_{t_{1i}i} = 1$. In the synthetic data used in this chapter, I assume no recruitment and simulate data with all animals present and seen on the first occasion.

4.4 Computation

Given a two-state hidden Markov model with underlying latent state chain, the probability of observation is denoted by π, and the transition probability matrix of the hidden Markov chain by γ. Given a series of observations $X_{1:T}$ and prior distributions on π and γ, our aim is to estimate the posterior

Algorithm 4.1 Pseudo algorithm for MCMC loop

The MCMC algorithm for one iteration consists of the following steps:

1. Sampling the hidden state chain for all individuals (see section 4.4.1).
2. Calculating summary statistics per individual conditional on its sampled states (see section 4.4.2).
3. Updating the posteriors for individual-level π_i,γ_i^{HH} and γ_i^{AA} separately using Gibbs sampling from Beta distributions (see section 4.4.3).
4. Updating the population-level hyper-parameters for π,γ^{HH} and γ^{AA} using an Independent Metropolis-Hastings sampler with three proposal distributions whose parameters vary across iterations (see section 4.4.4).
5. Updating population-level fixed effects using an Independent Metropolis-Hastings sampler with a fixed proposal distribution: a multivariate t-distribution whose mean and variance are set using a preliminary fit from ADMB (chapter 3) (see section 4.4.5).

distribution using MCMC. The MCMC routine developed in this chapter involves four main steps (five in application to real data); see Algorithm 4.1.

The R statistical software was used for model fitting, but a key component of the algorithm was implemented in C++, and called to R using the R-C interface. Step 1, the hidden Markov model Forward-Backward recursion (see next section), for all animals, is coded in C++. The remaining steps are all coded in R. Coding the hidden Markov model Forward-Backward recursion scheme and hidden state chain sampling into C++ greatly improved the speed of the model. Future development and further application of this work would involve coding each of the steps of the MCMC sampler into C++, or an alternative low level language.

4.4.1 Forward-Backward recursion

In order to update individual-level parameter values (θ_i for individual-level values and θ for population-level values) at each iteration, we require counts of successes and trials for each individual. Following Zucchini and MacDonald (2009) we simulate a sample path ($Z^{(T)}$) from the conditional distribution

$$\mathbb{P}(Z^{(T)}|x^{(T)},\theta) = \mathbb{P}(Z_T|x^{(T)},\theta) \times \prod_{t=1}^{T-1} \mathbb{P}(Z_t|x^{(T)}, Z_{t+1}^T, \theta)$$

We draw the sample path in the order $Z_T, Z_{T-1}, ..., Z_1$. To do this we need

$$\mathbb{P}(Z_t|x^{(t)},\theta) = \frac{\mathbb{P}(Z_t, x^{(t)}|\theta)}{\mathbb{P}(x^{(t)}|\theta)} = \frac{\alpha_t(Z_t)}{L_t} \propto \alpha_t(Z_t) \text{ , for } t = 1, ..., T.$$

where α_t are the forward probabilities. Given the forward probabilities we backward sample $(T, T-1, ..., 1)$ the hidden state chain. This recursion scheme is referred to as the Forward-Backward recursion scheme (Scott, 2002; Zucchini and MacDonald, 2009). The counts are then obtained from these sampled state chains. .

This recursion scheme consists of one forward pass and one backward pass, per individual, for each iteration of the MCMC sampler. The Forward recursion produces the forward probability vector $\alpha_2, ..., \alpha_n$, containing the probabilities of the underlying hidden states for each observation given all observed data up to time t. We calculate these forward probabilities, from $1:T$ (T being the length of the observation history), for each state, given the observed data (X):

$$\begin{aligned}\alpha_t(S_t) &= \mathbb{P}(S_t|X_{1:t}) \\ &= \sum_{S_{t-1}} \mathbb{P}(S_{t-1}|X_{1:t-1})\mathbb{P}(S_t|S_{t-1})\mathbb{P}(X_t|S_t) \\ &= \sum_{S_{t-1}} \alpha_{t-1}(S_{t-1})\mathbb{P}(S_t|S_{t-1})\mathbb{P}(X_t|S_t)\end{aligned}$$

where $\mathbb{P}(X_t|S_t)$ denotes the probability of the data given the state. The backward pass generates a sample path (Z) of the hidden state chain in the order $t = T, T-1, T-2, ..., 1$, making use of the following proportionality argument:

$$\mathbb{P}(Z_t|x^{(T)}, Z_{t+1}^T, \theta) \propto \alpha_t(Z_t)\mathbb{P}(Z_{t+1}|Z_t, \theta). \tag{4.1}$$

The second factor in equation 4.1 is simply a one-step transition probability in the Markov chain.

4.4.2 Calculating summary statistics per individual

Observations for an individual are Binomial with probability π_i and trial size the number of times they were recorded as Here ($S_t = 1$) in the sampled state chain in section 4.4.1. As the Beta prior for π is conjugate to the Binomial, the posterior is also Beta. In order to sample from this posterior, we require counts of the successes and trials for each individual at each iteration. These summary counts are obtained from either the data or the sampled state chains from the Forward-Backward recursion scheme described previously (section 4.4.1). For example, the counts of observations are obtained from the capture histories, and at each iteration the number of times Here is calculated from the sampled state chain from the Backward recursion. A

similar scheme applies to γ^{HH}: a hidden state of Here in the sampled state chain is Binomial with probability γ_i^{HH}, with trial size the number of times they remained Here (i.e. $Z_t = H$ and $Z_{t+1} = H$) or transitioned to away (i.e. $Z_t = H$ and $Z_{t+1} = A$). Given that the Beta prior for γ^{HH} is conjugate to the Binomial, the posterior for γ^{HH} is also Beta. Likewise for γ^{AA}.

For the probability of observation there is a trial whenever an animal is Here; the outcome is whether it was or wasn't seen. There is no trial when the animal is Away, since it is then guaranteed not to be seen. For γ^{HH}, there is again a trial whenever the animal was Here (excluding the final period); the outcome is whether it stayed Here or not. A similar scheme applies to γ^{AA}.

4.4.3 Updating individual values of γ^{HH}, γ^{AA} and π

Following the hidden Markov model Forward-Backward recursion scheme and calculation of the counts of successes and trials, we update individual-level random effects by sampling from the posterior distribution. Each individual-level parameter θ_i (π_i, γ_i^{HH} or γ_i^{AA}) is an independent sample from a prior Beta distribution: $\theta_i \sim \text{Beta}(a, b)$, where (a,b) will obviously be different for π, γ^{HH} and γ^{AA}.

The joint posterior is

$$\begin{aligned} p(\theta, a, b|y) &\propto p(a, b)p(\theta|a, b)p(y|\theta) \\ &\propto p(a, b)\prod_{i=1}^{N}\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\theta_i^{a-1}(1-\theta_i)^{b-1}\prod_{i=1}^{N}\theta_i^{y_i}(1-\theta_i)^{n_i-y_i} \end{aligned}$$

Gibbs sampling can be used to update θ_i, since the full conditional for θ is available: $\theta_i|a, b, y \sim Beta(y_i + a,\, n_i - y_i + b)$ where y_i indicates the

number of successes (e.g. number of observations) for individual i, and n_i the number of trials.

4.4.4 An Independent Metropolis-Hastings sampler - the choice of proposal distribution for a and b

Extending the hierarchy above to deal with the population-level hyper-parameters a and b we have

$$\begin{aligned} y_i|\theta_i &\sim Bin(\theta_i, n_i) \\ \theta_i|a,b &\sim Beta(a,b) \\ a,b &\sim p(a,b) \end{aligned}$$

where $p(a,b)$ indicates the prior distribution for the hyper-parameters. Given the posterior distribution of parameters is $p(\theta,a,b|y) \propto p(a,b)p(\theta|a,b)p(y|\theta)$ we can see that given θ, the dependency on the data disappears for the hyper-parameters. Thus in order to update our population-level hyper-parameters, we require only the updated values of θ. As there is no conjugate prior, these population-level hyper-parameters are updated using an Independent Metropolis-Hastings sampler.

The Independent Metropolis-Hastings sampler works by ignoring the current value θ^*, and sampling the candidate value for update, θ°, directly from a proposal distribution $\tilde{f}$ that should be close to the ideal distribution $f(\theta)$. Following this, the acceptance ratio becomes

$$\frac{f(\theta^*)}{f(\theta^\circ)} \frac{\tilde{f}(\theta^\circ)}{\tilde{f}(\theta^*)}$$

Algorithm 4.2 Moments of a logit-transformed Beta distribution. Algorithm by M.V. Bravington (pers comm).

The moments are easily derived from differentiating the Cumulant Generating Function

$$\begin{aligned}
\mathbb{E}\left[\exp\left(t\text{logit}p\right)\right] &= \frac{1}{B\left(a,b\right)}\int \exp\left(t\log p - t\log\left(1-p\right)\right)p^{a-1}\left(1-p\right)^{b-1}dp \\
&= \frac{1}{B\left(a,b\right)}\int p^{a+t-1}\left(1-p\right)^{b-t+1}dp \\
&= \frac{B\left(a+t,b-t\right)}{B\left(a,b\right)} \\
\Longrightarrow\quad & K\left(t\right) = \log\mathbb{E}\left[\exp\left(t\text{logit}p\right)\right] \\
= \quad & \log\Gamma\left(a+t\right)+\log\Gamma\left(b-t\right)-\log\Gamma\left(a+b\right)-\log\Gamma\left(a\right)-\log\Gamma\left(b\right)+\log\Gamma\left(a+b\right) \\
\Longrightarrow\quad & K'\left(t\right) = \psi\left(a+t\right)-\psi\left(b-t\right) \\
& K''\left(t\right) = \psi'\left(a+t\right)+\psi'\left(b-t\right) \\
\Longrightarrow\quad & \mu = \mathbb{E}\left[\text{logit}p\right] = K'\left(0\right) = \psi\left(a\right)-\psi\left(b\right) \\
& \sigma^2 = \mathbb{V}\left[\text{logit}p\right] = K''\left(0\right) = \psi'\left(a\right)+\psi'\left(b\right)
\end{aligned}$$

$$\text{where } \psi(x) = \frac{d}{dx}log\Gamma(x)$$

which does not completely cancel. However, insofar as the approximating distribution is close to the target distribution, the average acceptance ratio will be close to 1.

The logit scale is used, since the distribution of logitθ is reasonably Normal for any reasonable Beta prior on θ (i.e. unless (a,b) have become extreme). The collection of logitθ is distributed approximately Normal, $N(\mu,\sigma^2)$. We need to update this distribution with the collection of individual θ_i to get a posterior for logitθ - $N(\mu',\sigma^{2\prime})$. In order to do this we need to turn $(\mu',\sigma^{2\prime})$ into the corresponding (a,b). There is a simple relationship between (a,b) and $\left(\mu,\sigma^2\right)$, the mean and variance of logitθ (see Algorithm 4.2). To get (a,b) from (μ,σ^2) we apply a Newton-Raphson iteration to

$$\begin{aligned} \mu &= \psi(a) - \psi(b) \\ \sigma^2 &= \psi'(a) + \psi'(b). \end{aligned}$$

Given approximate starting values, $\mu = ln(a)/ln(b)$ and $\sigma^2 = ln(a+b)$, we apply a Newton-Raphson iteration to solve for (a, b). For a given $\left(\mu, \sigma^2\right)$, we seek

$$(a,b) \quad \text{s.t.} \quad F(\theta) = \begin{bmatrix} \psi(a) - \psi(b) \\ \psi'(a) + \psi'(b) \end{bmatrix} - \begin{bmatrix} \mu \\ \sigma^2 \end{bmatrix} = 0.$$

Thus

$$F'(\theta) = \begin{bmatrix} \psi'(a) & -\psi'(b) \\ \psi''(a) & \psi''(b) \end{bmatrix}$$

and we update via $\theta^{r+1} = \theta^r - \left[F'(\theta^r)\right]^{-1} F(\theta^r)$, where $r+1$ indicates the next iteration.

For the approximating distribution $(\tilde{f})$, we assume a vague conjugate-prior for the mean and variance $\left(\mu, \sigma^2\right)$ with the following conjugate hyper-priors for the logitθ_i

$$\begin{aligned} \text{logit}\theta_i \quad |\mu, \tau \quad &\sim N(\mu, \tau) \\ \mu \quad |\tau \quad &\sim N(\mu_0, n_0\tau) \\ \tau \quad &\sim Ga(\alpha_\tau, \beta_\tau) \end{aligned}$$

where $\mu_0 = 0$, $n_0\tau = 0.1$, $\alpha_\tau = 0.1$ and $\beta_\tau = 0.1$. Given the collection of logitθ_i we update this to a conjugate posterior $\tilde{f}\left(\mu, \sigma^2 | \text{logit}\theta_{\mathcal{I}}\right)$ in the

standard way for conjugate Gaussian problems. Following this we sample $\theta^* = \left(\mu^*, \sigma^{2*}\right)$ from $\tilde{f}(.)$ and compute $\tilde{f}(\theta^*)$; then $\left(\mu^*, \sigma^{2*}\right)$ are back-transformed to (a^*, b^*). Following this we compute the (vague prior times) log-likelihood of the collection of $\theta_{\mathcal{I}}$ under (a^*, b^*) and current values; and finally the acceptance ratio is calculated and the hyper-parameters are updated accordingly.

4.4.5 Updates to fixed effects

Although not used in simulations, the inclusion of any fixed effects (one population-level value, no random effects) in the model are updated using an unchanging Independent Metropolis-Hastings sampler. For example: for individual i, given a population-level fixed effect b, individual-level effect α_i, and a design matrix X_i, the individual-level parameter of interest ψ_i, is calculated as $\text{logit}\psi_i = \alpha_i + X_i b$. If α_i are sampled from a Beta distribution then these individual-level effects are first transformed using the logit-link.

An Independent Metropolis-Hastings sampler is used to update these values as there is no conjugate prior and, unlike in section 4.4.4, it is not obvious how to generate an approximate adaptive proposal. The use of an Independent Metropolis-Hastings sampler avoids the need to adjust the tuning parameters required in a random walk MCMC and the Laplace approximation provides a good approximation to the posterior for these parameters which can be used to design the unchanging proposal distribution.

The update to population-level parameters can be made using a multivariate t-distribution with 5 degrees of freedom (MVT_5) with covariance structure based on results from ADMB. The same structure and data was used for the model in ADMB and the results were used to facilitate block updates from the multivariate t-distribution. In comparison to a multivariate

Normal proposal, the multivariate t-distribution allows for the potential of thick tails in the posterior and better mixing of the chain. The use of a multivariate distribution is more efficient than block updates of univariate distributions when there is strong correlation between parameters and results in an efficient proposal distribution which is close to the expected posterior estimates. The log-likelihood is used for computations in the Metropolis-Hastings acceptance ratio to avoid potential numerical instabilities which may occur with use of the likelihood.

4.5 Simulation and convergence testing

4.5.1 Convergence with one long chain

A large simulated data set for 60 animals with capture history of length 500 was used as an exploratory test of the model. A MCMC chain was run for 100000 iterations, with the first 10000 discarded to burn in. Heterogeneity was included on each of the two states and on the probability of observation. Individual probabilities were drawn from the following beta distributions: $\pi \sim Beta(4,2)$, $\gamma^{HH} \sim Beta(6,2)$, $\gamma^{AA} \sim Beta(15,5)$.

Trace plots (Figure 4.1) and density plots (Figure 4.2) for the six hyper-parameters indicate good mixing and convergence. These parameters were updated using the Independent Metropolis-Hastings algorithm (see section 4.4.4); the mean acceptance rate was 80%. True values plotted against the mean posterior for each individual for π, γ^{HH} and γ^{AA} (Figure 4.3) indicate strong correlation between mean of individual posterior estimates and the true values used in simulation.

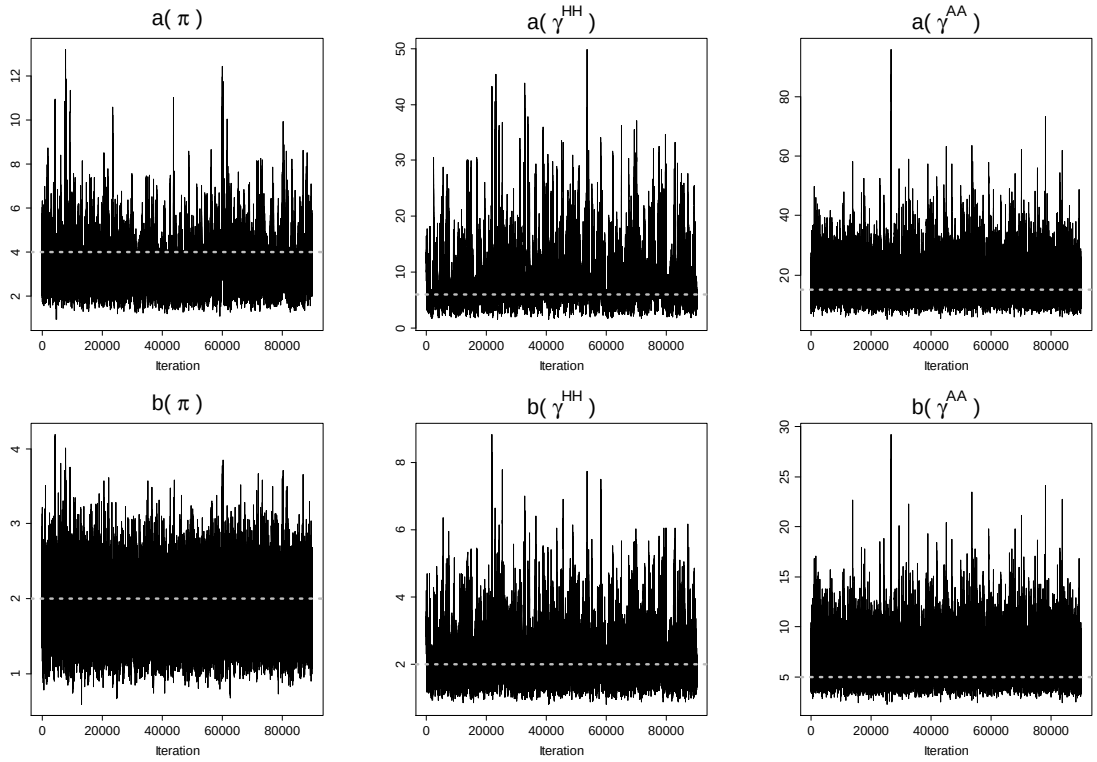

Figure 4.1: Trace plots for population-level hyper-parameters. Dotted grey lines indicate true values used in simulation of synthetic data. Plots indicate convergence of chains.

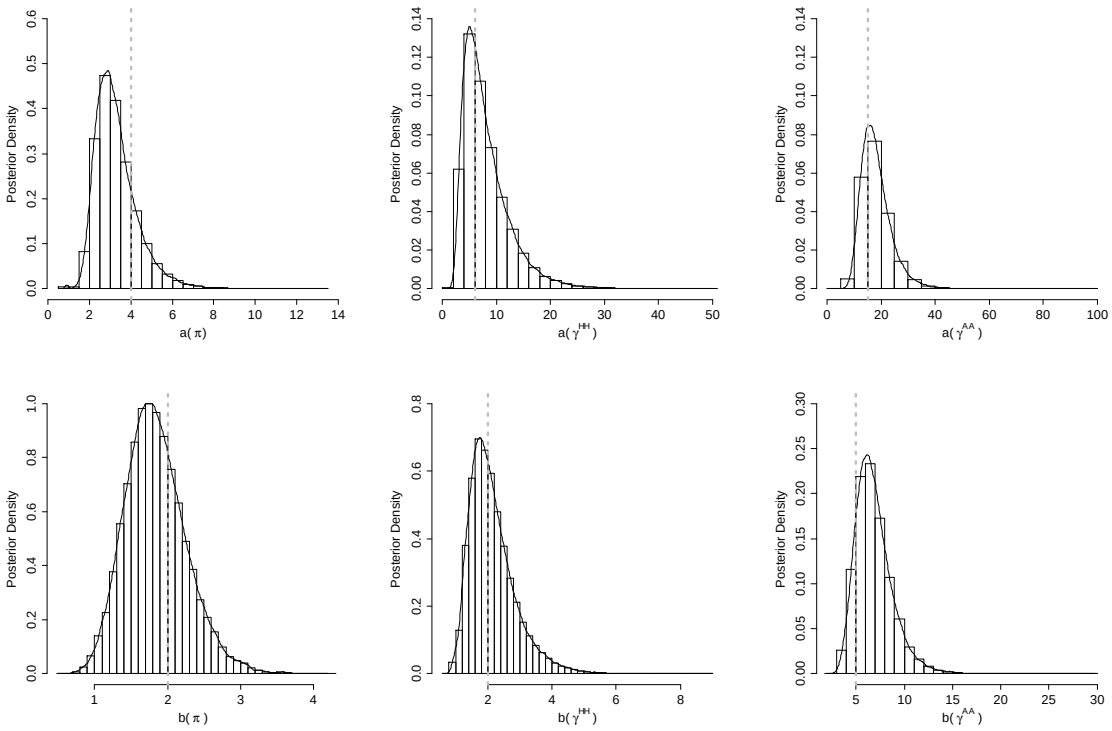

Figure 4.2: Histogram and density plot of population-level hyper-parameters. Vertical dashed lines indicate true value used to simulate data.

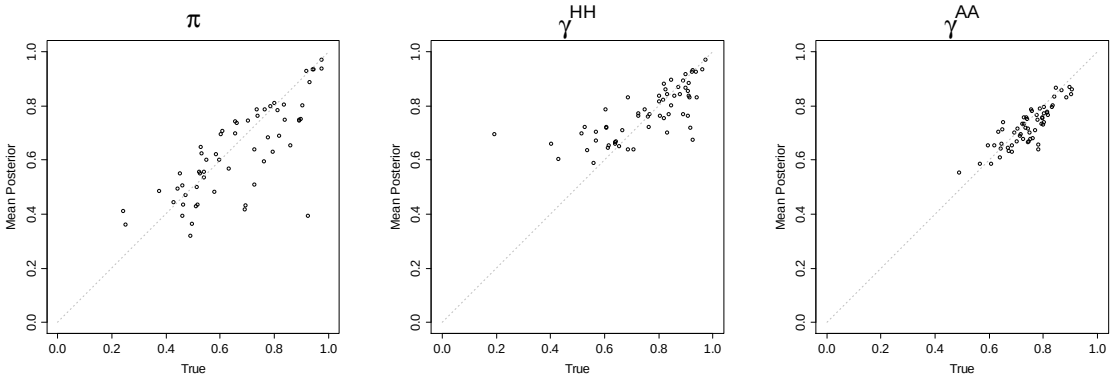

Figure 4.3: Plot of true values from simulated data vs mean of individual posterior values for π, γ^{HH} and γ^{AA}.

	π		γ^{HH}		γ^{AA}	
	a	b	a	b	a	b
TRUTH	8	2	8	2	8	2
50 animals	13.17 (**6.41**)	2.71 (**1.00**)	7.34 (**2.14**)	1.93 (**0.45**)	10.40 (**2.60**)	2.33 (**0.53**)
200 animals	9.47 (**1.64**)	2.29 (**0.34**)	7.07 (**1.17**)	1.77 (**0.24**)	6.60 (**0.94**)	1.65 (**0.19**)
800 animals	7.98 (**0.88**)	1.97 (**0.17**)	8.62 (**0.93**)	2.17 (**0.17**)	8.77 (**0.68**)	2.08 (**0.13**)

Table 4.1: Means and standard deviations (in brackets) for hyper-parameters.

4.5.2 Asymptotic convergence of a and b

The asymptotic convergence of the posterior distribution for each of the hyper-parameters was tested using three simulated data sets with fixed capture histories (length 250) but increasing number of animals (50, 200, 800). For each parameter (π, γ^{HH} and γ^{AA}), synthetic data was simulated from a $Beta(8, 2)$ distribution (i.e. a mean of approximately 0.8). MCMC chains were run for 15000 iterations, with the first 5000 discarded to burn in.

Figure 4.4 shows density plots of posterior estimates for the six hyper-parameters: decreasing variance is evident with increasing number of individuals. Table 4.1 indicates posterior standard deviations decreased at the expected rate of approximately $\frac{1}{\sqrt{n}}$.

4.5.3 Convergence of individual mean posterior estimates

Asymptotic convergence for the mean of individual posterior estimates to true individual values (used in simulation) was tested by increasing length of capture history (250, 1000, 4000) for three data sets each with 16 individuals. MCMC chains were run for 15000 iterations, with the first 5000 discarded to burn in. Data were simulated using the following distributions: $\pi \sim Beta(30, 3)$, $\gamma^{HH} \sim Beta(30, 5)$ and $\gamma^{AA} \sim Beta(30, 2)$ (i.e. approximate mean probabilities of 0.91, 0.86, 0.94).

Figure 4.5 indicates convergence of the mean of individual posterior es-

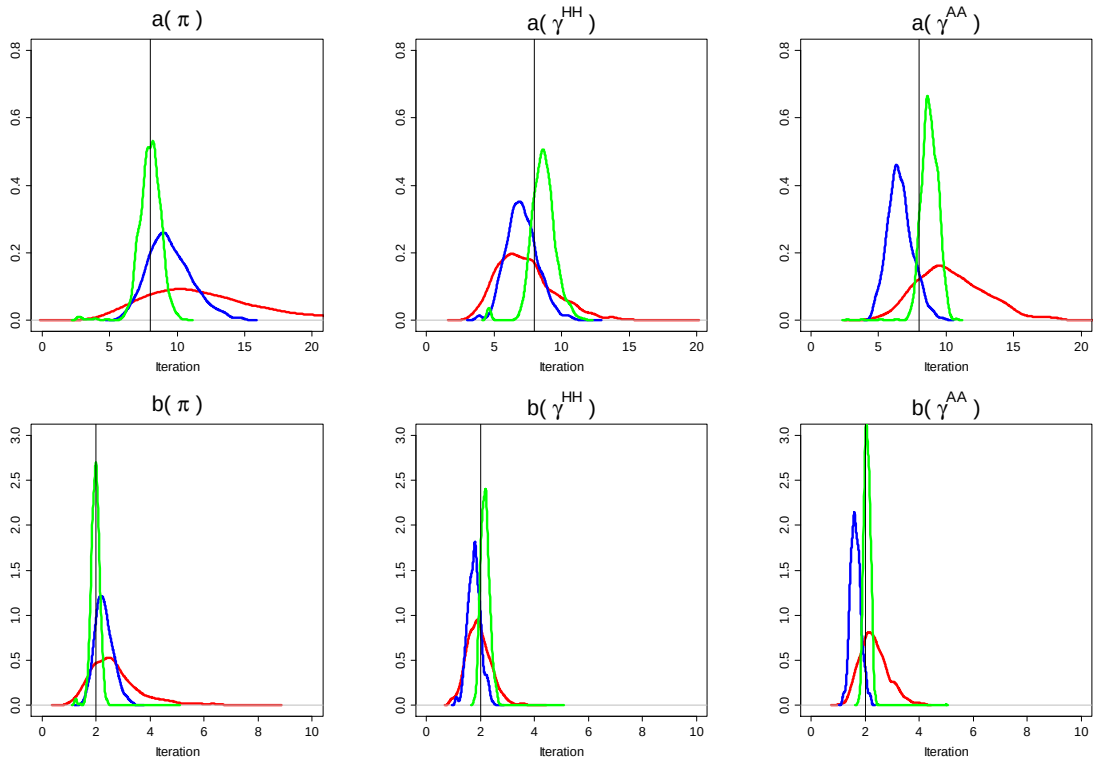

Figure 4.4: Posterior density plots for each hyper-parameter. The red density line is for 50 animals; blue for 200 animals; and green for 800 animals. The vertical black line indicates the true values used to simulate the data.

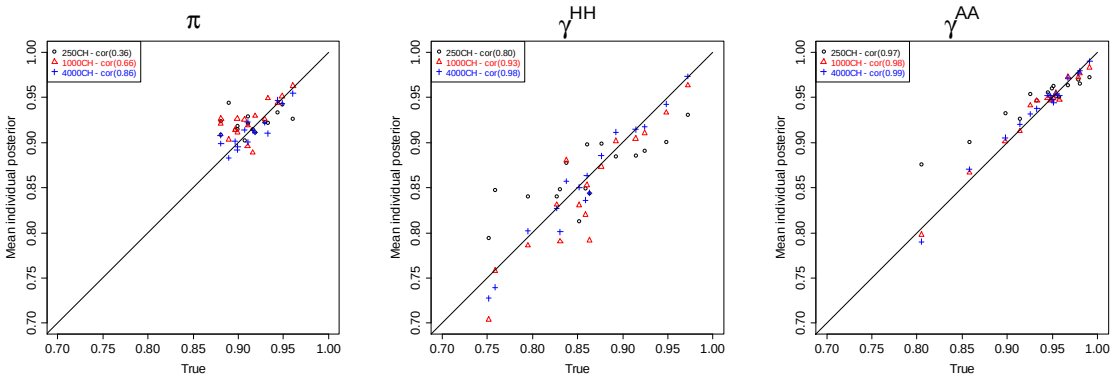

Figure 4.5: Mean individual posterior values vs true values for π, γ^{HH} and γ^{AA}. Correlations indicate increasing convergence to true values with increasing length of capture history.

timates to true values for each of π, γ^{HH} and γ^{AA}. Standard deviations decreased with increasing length of capture history, and correlation improved with longer capture histories.

4.6 Application to North Atlantic Humpback whales

The full three-state hidden Markov model including death, developed in chapter 3, was applied to a sub-set of 176 North Atlantic Humpback whales sighted more than 30 times in the Stellwagen Bank National Marine Sanctuary (SBNMS) between 1979 and 2005. A subset of the real data was used, rather than all 1147 whales as per chapter 3, due to the computational speed of the MCMC sampler. Individual heterogeneity was included on: probability of observation, probability of remaining here, and probability of remaining away. For clarity several sections from chapter 3 are briefly reiterated here.

With three states the nine elements of the transition matrix can be written in terms of just three parameters γ^{HH}, γ^{AA}, and γ^{D} (respectively the

probabilities of staying Here, staying Away and Dying) as follows:

$$\gamma = \begin{pmatrix} \gamma^{HH}\left(1-\gamma^{D}\right) & \left(1-\gamma^{HH}\right)\left(1-\gamma^{D}\right) & \gamma^{D} \\ \left(1-\gamma^{AA}\right)\left(1-\gamma^{D}\right) & \gamma^{AA}\left(1-\gamma^{D}\right) & \gamma^{D} \\ 0 & 0 & 1 \end{pmatrix}$$

where it is assumed that the probability of death (which is very low relative to the other transition rates) does not depend on whether the animal is Here or Away.

Only data from the first sighting of each whale onwards are used. As sighting effort is focused in the middle of the year, we included all sightings from the 18th week of the year through to the 43rd week. The probability of survival, P_{surv}, over the remaining 26 week period was calculated as $P_{surv} = (1-\gamma^{D})^{26}$. An extra parameter q was introduced for the chance of being present in the SBNMS at the start of the season. We calculated the probability of each state in the first week of the new year to be:

$$\mathbb{P}(S_t|S_{t-1}) = \begin{pmatrix} q * P_{surv} & (1-q) * P_{surv} & 1-P_{surv} \\ q * P_{surv} & (1-q) * P_{surv} & 1-P_{surv} \\ 0 & 0 & 1 \end{pmatrix} * \mathbb{P}(S_{t-1})$$

where $\mathbb{P}(S_{t-1})$ is the vector of state probabilities in the last week of the previous year.

Individual-level random effects were included on each of π, γ^{HH} and γ^{AA} and are updated exactly as described for the two-state model applied to synthetic data (see sections 4.4.1, 4.4.2, 4.4.3, 4.4.4). We assume individual-level parameters to be consistent over time but have allowed for population-

level annual variation in probability of remaining Here using logit-links: logit$\gamma_{i,yr}^{HH} = \beta_{yr} + \gamma_i^{HH}$. Population-level parameters -β_{yr}, γ^D (death) and q - are updated as in section 4.4.5.

4.6.1 Results for North Atlantic Humpback whales

One chain was run for 100000 iterations with the first 10000 discarded to burn-in. Figure 4.6 indicates the trace plots for the first and last 45000 posterior estimates and Figure 4.7 shows density plots of the mean and standard deviation of logit π, γ^{HH} and γ^{AA} (calculated as $mean = \psi(a) - \psi(b)$ and $sd = \sqrt{\psi(a)' + \psi(b)'}$). Both the trace plots and a transform to the logit scale indicate no visible differences between the first and second half of the chains suggesting that full convergence was reached with 100000 iterations.

Posterior estimates were less varied for π compared to γ^{HH} and γ^{AA}. Figure 4.8 indicates the density for π, γ^{HH} and γ^{AA} for all individuals for 900000 iterations. Population-level variability is apparent in π, but somewhat less so in the two state probabilities. Since the transition parameters are close to 1, there is less opportunity to see heterogeneity than for π. However, Figure 4.7 indicates that the standard deviation logits are not much smaller for both γ^{HH} and γ^{AA} compared to π.

4.6.2 Comparison with ADMB

We compared the performance of this method against the Laplace approximation using ADMB-RE. Results from the full model with the same subset of 176 whales was run in ADMB. Figure 4.9 shows results for individual posterior estimates for π and γ^{AA} from both ADMB and from the MCMC sampler developed here. Results appear to be similar, but more variation

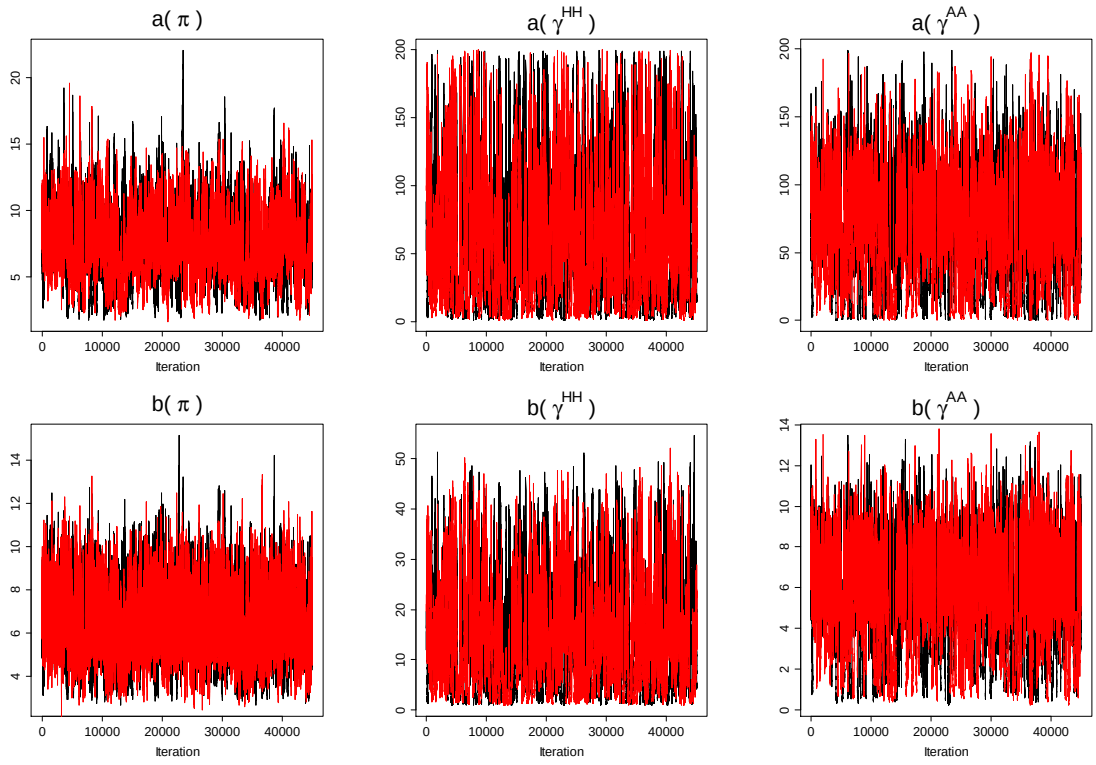

Figure 4.6: Trace plots of hyper-parameters for π, γ^{HH} and γ^{AA} for 90000 posterior estimates. The black trace indicates the first 45000 estimates and the red the final 45000.

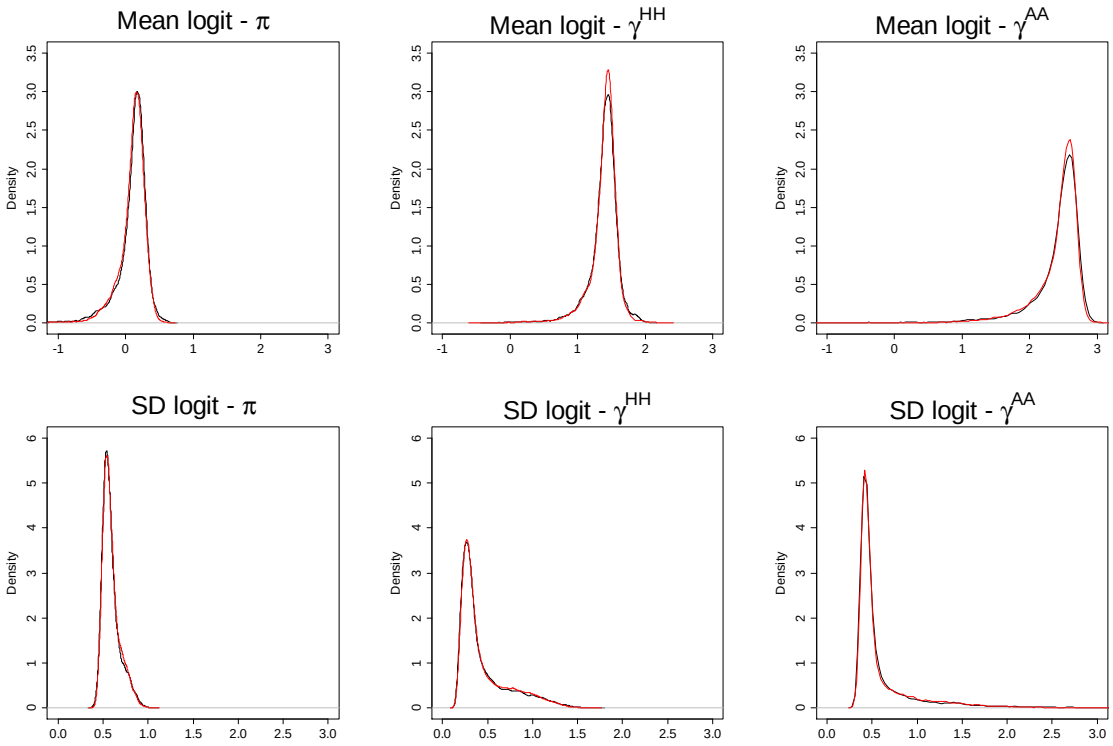

Figure 4.7: Density plots for the mean and standard deviation of logit π, γ^{HH} and γ^{AA} for 90000 posterior estimates. The black line indicates the first 45000 estimates and the red the final 45000.

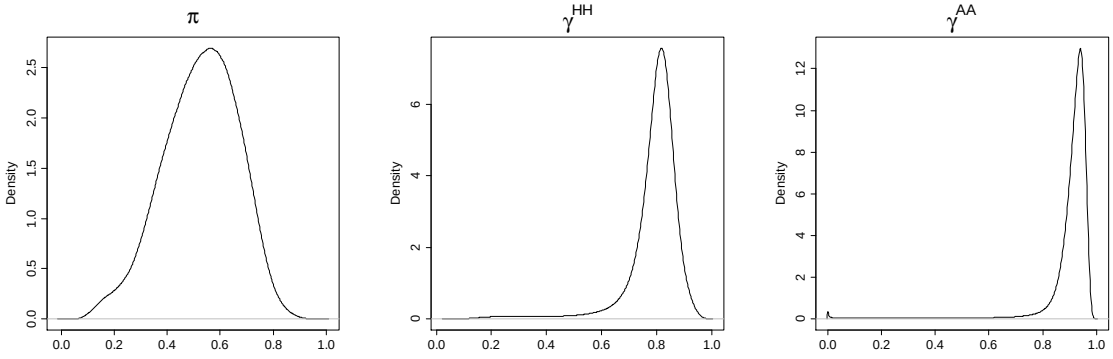

Figure 4.8: Density plots from 90000 posterior samples from three independent chains for π, γ^{HH} and γ^{AA}.

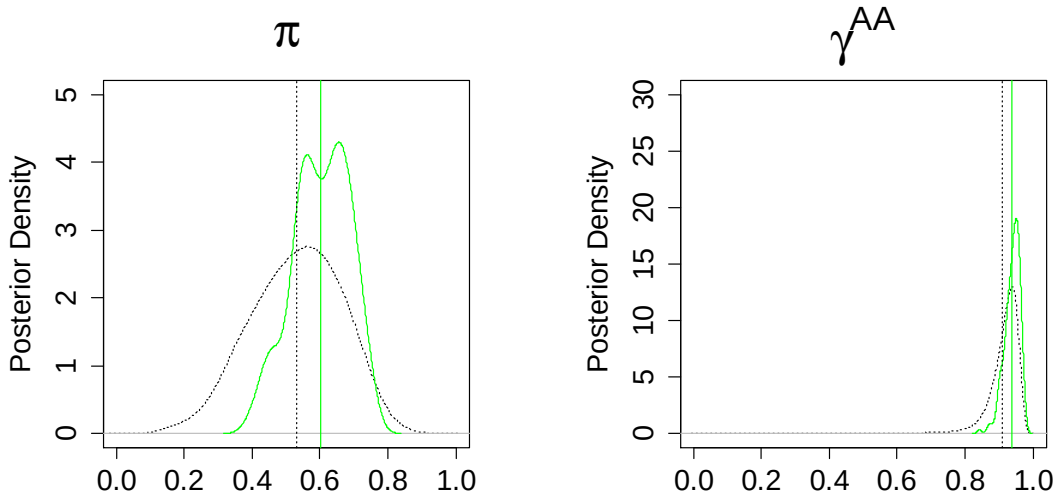

Figure 4.9: The dashed black line indicates density from 6000 posterior samples from three independent chains for π and γ^{AA} and the green line individual posterior estimates from ADMB. Vertical lines indicate means of posterior estimates.

and slightly lower mean is evident in the Beta-Binomial results. This difference in variation is likely due to the Beta-Binomial results incorporating the uncertainty on the hyper-parameters; in comparison ADMB results are conditioned on point estimates of the hyper-parameters.

4.6.3 Exploration of datasets with no heterogeneity

One limitation we found in application to smaller subsets of real data (not presented above), was that in some cases the Independent Metropolis-Hastings sampler got stuck (on updates to the state transition probabilities), resulting in low acceptance rate of proposals (~20-30%). This was likely due to lack of variability between individuals. This was verified from exploration of simulation studies: trace plots for hyper-parameters for datasets with no heterogeneity (i.e. synthetic data simulated using discrete values for parameters)

revealed lower acceptance rate, and often much higher posterior estimates of hyper-parameters, in comparison to data sets with heterogeneity. However, the high posterior estimates on one or both of the hyper-parameters (e.g. $a = 1500$, $b = 80$), indicated that the MCMC had in fact converged to a distribution which reflected the true value used to simulate the synthetic data.

The low acceptance rate in application to smaller subsets of real data may have also in part been due to the uninformative vague hyper-prior used on updates to population-level hyper-parameters a and b (see section 4.4.4). Some proposals resulted in a distribution of individual values with a similar mean, but a larger variance. This increased variance resulted in biologically unreasonable responses as animals were allowed both low γ^{HH} and γ^{AA}. If both γ^{HH} and γ^{AA} are below 0.5, for example, then it would imply that animal i alternates between Here and Away in successive weeks. One potential permanent fix to this would be to make modifications to the joint prior for γ^{HH} and γ^{AA}, in order to allow only one of these two state probabilities to be below, say 0.6. This method would penalize the proposed updates to a and b if they resulted in too many low individual probabilities, for example more than half of the values below 0.6.

4.7 Discussion

The methods developed in this chapter provide an alternative approach to the Laplace approximation in chapter 3 and outline the MCMC sampling framework required for the more complex non-parametric approach developed in the next chapter. We have developed a hierarchical Bayesian hidden Markov model where individual parameters π, γ^{HH} and γ^{AA} are updated

using Gibbs sampling, the population-level hyper-parameters using a changing Independent Metropolis-Hastings sampler and fixed effects using an unchanging Independent Metropolis-Hastings sampler.

Results from simulation tests indicate asymptotic convergence for both individual and population-level parameters: the posterior standard deviations for the population-level parameters decreased with increasing number of individuals, and correlation for mean posterior individual estimates to true values improved with increasing length of capture history.

Output from ADMB was used to facilitate sampling from a multivariate t-distribution in order to update the fixed parameters in the model. This novel method resulted in an automatic and computationally efficient method to update these complex model parameters, avoiding the need to tweak, for example, random walk step sizes. This method could also be used to incorporate covariates into the model, providing another extension to the already flexible model.

A limitation of both ADMB, with a Normal prior, and the Beta-Binomial method is the assumption of unimodality. We extend this model in the next chapter to incorporate the non-parametric Dirichlet process prior which allows for unimodal or multimodal behaviour in each of the three parameters that describe individual-level heterogeneity.

Chapter 5

Modeling latent individual heterogeneity with Dirichlet process priors.

5.1 Introduction

In mark-recapture literature, modeling individual heterogeneity most often involves either the use of a preset functional form (e.g. Gaussian), or assignment to a prespecified number of groups. The use of any preset form is limited and limiting, as it enforces strict assumptions on the expected distribution of the population. Having to make assumptions about the number of groups *a priori* can result in model selection problems (Cubaynes et al., 2012). The natural subgroups often seen in mark-recapture studies and the complexity of real mark-recapture data means that parametric and discrete style models can be insufficient. Non-parametric models avoid these often restrictive assumptions. In this chapter we consider the non-parametric Dirichlet process for modeling latent individual heterogeneity. The Dirich-

let process prior is a flexible extension to a parametric model as it avoids assumptions about the functional form of the distribution, and it extends discrete style models to the infinite limit by avoiding any prespecifications about the number of groups (Dorazio et al., 2008; Navarro et al., 2006).

Despite the appeal of the Dirichlet process prior, it has had little application in mark-recapture analysis. This is most likely due to its complexity and somewhat confusing literature. One good example of the Dirichlet process applied to mark-recapture data, is the paper by Dorazio et al. (2008), who used it to model animal abundance where heterogeneity in abundance between sites was poorly understood and not directly observable.

The Dirichlet process is introduced in section 5.2. The remainder of the chapter follows a similar course to chapter 4. The difference in this chapter is that individual-level values are updated using a Dirichlet process prior. This involves several steps which are outlined in section 5.3. Following this in section 5.4, synthetic data is used to test the model. The full three-state model (including death, using the same structure as chapter 3 and 4) is applied to North Atlantic humpback whales in section 5.5.

5.2 The Dirichlet process prior

The Dirichlet process was first introduced by Thomas Ferguson in 1973 (Ferguson, 1973). It is a stochastic process defined as a distribution on distributions and is defined by two parameters: the base distribution G_0, and the precision parameter α. As the Dirichlet process supports all distributions, one might expect the output from a Dirichlet process to be continuous. Although the base distribution may be continuous, individual draws G from the Dirichlet process are discrete with probability one (Blackwell and Mac-

Queen, 1973; Ferguson, 1973; Neal, 2000; Sethuraman, 1994). This means that draws from a Dirichlet process will be clustered on a countably infinite set of discrete values. The result is that values will be repeated, as individuals in the same cluster will have the same value. The precision parameter guides this clustering: the lower α is, the more variability will be observed between individual realisations, and for any given realisation a small α will correspond to a smaller number of clusters (see Figure 5.1). The influence of α on the number of clusters can be seen in Figure 5.1, with the number of clusters increasing with α, along with the concentration of draws around G_0 for large α. The number of clusters will tend to ∞ with high values of α; conversely the number of clusters will tend to 1 with low values of α. In comparison to this non-parametric approach, finite mixtures must specify the number of clusters *a priori*. As such, the Dirichlet process is the infinite limit of the discrete groups approach which assumes a fixed number of groups. In this way α corresponds to the strength of prior belief in the base distribution and the number of clusters which are likely to be sampled from it. Note that G_0 itself will generally be of specified parametric form, e.g. Normal, and will have unknown parameters which are updated separately to the Dirichlet process.

A generic Dirichlet process takes the form

$$\begin{aligned} y_i|\theta_i &\sim F_i(\theta_i) \\ \theta_i|G &\sim G \\ G &\sim DP(G_0, \alpha). \end{aligned}$$

Here, we assume that data y_i are independent conditional on θ_i, and G is the mixing distribution over θ which has Dirichlet process prior $DP(G_0, \alpha)$.

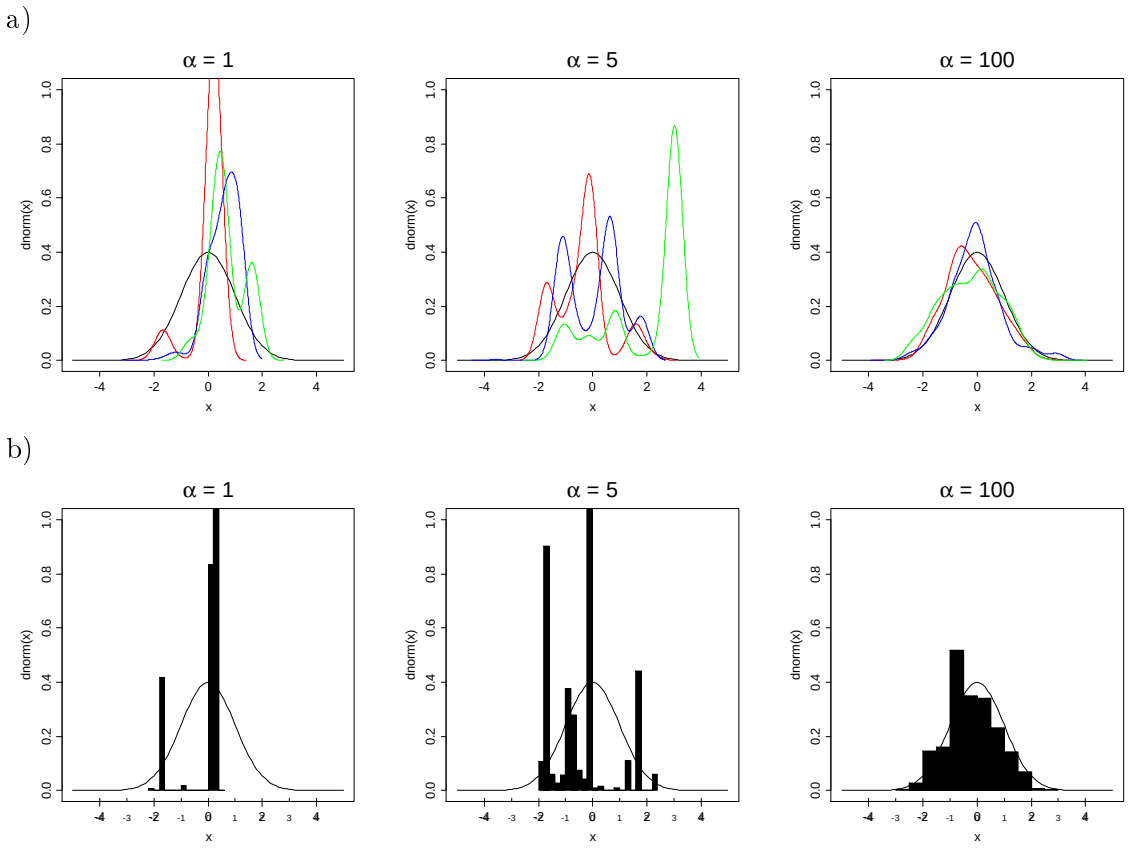

Figure 5.1: Simulations from a Chinese restaurant process for different weights of $\alpha(1, 5, 25, 100)$, $N = 10000$ and $G_0 \sim N(0, 1)$. (a) smoothed density curves (green, red and blue) for three independent realisations from a Chinese restaurant process for differing values of α. The black density line indicates the base distribution G_0. (b) histograms for the discrete draws that make up one individual realisation. The black density line indicates the base distribution G_0.

Several well known methods for the representation of a Dirichlet process include the Polya urn scheme (Blackwell and MacQueen, 1973) or Chinese restaurant process (Pitman, 2006) and the stick-breaking prior (Sethuraman, 1994; Ishwaran and James, 2001). Following is a description of the Chinese restaurant process which is the basis of the algorithm used in this chapter.

Consider the analogy of a Chinese restaurant with infinite seating capacity. The first customer enters the restaurant and sits at table one with probability one. Each subsequent customer entering the restaurant chooses a table with probability proportional to the number of people already seated at the table, or a new table proportional to α. Customers at the same table are served the same dish; customers at new tables are served a new dish at random. In this sense, individuals at each table receive the same parameter value (dish), and the table number indicates their cluster membership. In general terms, this means that the probability of seeing an already seen cluster is proportional to the number of individuals in that cluster, and the probability of seeing a new cluster is proportional to the precision parameter α (see Figure 5.1).

MCMC algorithms are the most common approach for inference in Dirichlet processes. Neal (2000) presented several algorithms which use the Chinese restaurant process approach to sample from the posterior distribution of the Dirichlet process. This chapter follows the work by Neal (2000). Alternative samplers include: blocked Gibbs sampling using the stick-breaking representation (Ishwaran and James, 2001); updates using a Metropolis-Hastings framework (Jain and Neal, 2004; Liang et al., 2007); and sequential Monte Carlo (Fearnhead, 2004).

Algorithm 5.1 Algorithm 8 by Neal (2000)

Let the state of the Markov chain consist of $c = (c_1, ..., c_n)$ and $\phi = (\phi_c : c \in \{c_1, ..., c_n\})$. Repeatedly sample as follows:

- For $i = 1, ..., n$ [where n indicates the number of individuals]: Let k^- be the number of distinct c_j for $j \neq i$, and let $h = k^- + m$. Label these c_j with values in $\{1, ..., k^-\}$. If $c_i = c_j$ for some $j \neq i$, draw values independently from G_0 for those ϕ_c for which $k^- < c \leq h$. If $c_i \neq c_j$ for all $j \neq i$, let c_i have the label $k^- + 1$, and draw values independently from G_0 for those ϕ_c for which $k^- + 1 < c \leq h$. Draw a new value for c_i from $\{1, ..., h\}$ using the following probabilities:

$$P(c_i = c | c_{-i}, y_i, \phi_1, ...\phi_h) \quad = \quad \begin{cases} b\frac{n_{-i,c}}{n-1+\alpha} F(y_i, \phi_c) \text{ for } 1 \leq c \leq k^- \\ b\frac{\alpha/m}{n-1+\alpha} F(y_i, \phi_c) \text{ for } k^- < c \leq h \end{cases}$$

  where $n_{-i,c}$ is the number of c_j for $j \neq i$ that are equal to c, and b is the appropriate normalizing constant. Change the state to contain only those ϕ_c that are now associated with one or more observations.

- For all $c \in \{c_1, ..., c_n\}$: Draw a new value from $\phi_c \,|\, y_i$ such that $c_i = c$, or perform some other update to ϕ_c that leaves this distribution invariant.

5.2.1 Gibbs sampling via the Dirichlet process prior

In comparison to previous chapters, in this chapter, the individual-level random effects (π_i, γ_i^{HH} and γ_i^{AA}) are updated separately using a Dirichlet process prior which follows algorithm 8 by Neal (2000) (see algorithm 5.1).

The algorithms in Neal's paper (2000) work by assigning individuals to clusters. Due to the clustering property of the Dirichlet process, some of the individual parameter values θ_i will be identical, and each θ_i is associated with a cluster. Indicator variables c_i are used to indicate the current cluster membership for each individual (which may change over the course of the MCMC) and the clustering of individuals means that the number of active clusters will typically be much smaller than N; K is used to refer to

the number of active clusters. For $k = 1, ..., K$, each cluster c_k will have associated parameter value ϕ_k.

Algorithm 8 in Neal's paper (2000) was chosen as it is a flexible extension to each of the previous algorithms in his paper, and it allows for efficient Gibbs sampling with a non-conjugate distribution. At each iteration, the algorithm temporarily includes m auxiliary components; these are new potential values for clusters, which may or may not actually get individuals assigned to them. For each individual, when updating c_i, either an existing cluster is chosen or one of these m new components. The probability of joining an existing cluster will be proportional to the number of individuals in that cluster, and the probability of joining a new cluster will be proportional to α/m, the prior precision split equally among the auxiliary components. These auxiliary components are generated i.i.d from the base distribution and are discarded at each iteration if not used by the Gibbs sampler (i.e not chosen as a new cluster). The use of auxiliary components avoids the need to integrate with respect to the distribution G_0 as these auxiliary components represent the new possible components. This approach is similar to methods developed by MacEachern and Muller (1998) in that auxiliary components are used to update the model, with the difference that the auxiliary components exist only temporarily in Neal's algorithm.

Following algorithm 8 in Neal's paper (2000) (see algorithm 5.1), individual parameter values π_i , γ_i^{HH} or γ_i^{AA} are updated by generating and assigning new clusters. In algorithm 5.1, $F(y_i, \phi_c)$ is calculated as the density under a Binomial and c_i indicates which latent class is associated with data y_i (here the counts of successes and trials for each individual), where the numbering of c_i is of no significance. For each class, c, the parameter ϕ_c determines the associated probability for that class; the collection of all ϕ_c

Algorithm 5.2 Pseudo algorithm for MCMC loop

The MCMC algorithm for one iteration consists of the following steps:

1. Sampling the hidden state chain for all individuals (see section 4.1 in chapter 4).
2. Calculating summary statistics per individual conditional on its sampled states (see section 4.2 in chapter 4).
3. Updating the posteriors for individual-level π_i, γ_i^{HH} and γ_i^{AA} separately using Gibbs sampling from the Dirichlet process prior (see section 5.2.1).
4. Updating the base distribution and precision parameter:
   (a) Updating the base distribution G_0 using an Independent Metropolis-Hastings sampler with three proposal distributions whose parameters vary across iterations (see section 5.3.1).
   (b) Updating the precision parameter (see section 5.3.2).
5. Updating population-level fixed effects using an Independent Metropolis-Hastings sampler with a fixed proposal distribution: a multivariate t-distribution whose mean and variance are set using a preliminary fit from ADMB (see section 4.5 in chapter 4).

is denoted by ϕ.

5.3 Estimation

Steps 1, 2 and 5 in algorithm 5.2 are the same as in chapter 4. Step 3 and 4 differ from chapter 4 as individual-level parameter values are updated using a Dirichlet process prior (outlined in section 5.2.1). Following this, the hyper-parameters for the base distribution G_0 are updated using the machinery of the Independent Metropolis-Hasting sampler developed in chapter 4 (see section 5.3.1) and the precision parameter α is updated as described in section 5.3.2. The population-level fixed effects are updated as per chapter 4.

5.3.1 Updating hyper-parameters (a, b) for the base distribution G_0

In order to update the hyper-parameters governing the base distribution G_0, it suffices to use the machinery for the Independent Metropolis-Hastings sampler developed in chapter 4. The updating occurs differently in that the values passed to the Independent Metropolis-Hastings sampler are the parameter values for each cluster (rather than individual parameter values as in chapter 4).

5.3.2 Updating the precision parameter α

Despite its importance, there is a lack of agreement in the literature outlining efficient methods to update the precision parameter (Dorazio, 2009; Kyung et al., 2010; Navarro et al., 2006; Escobar and West, 1995). A Gamma(a, b) prior is commonly used due to its conditional conjugacy. However, the problem is knowing how to efficiently update the Gamma hyper-parameters (a, b).

The most recent and concise work in this field is by Murugiah and Sweeting (2012) who propose values for the hyper-parameters which can be used in the presence or absence of information. They suggest that standard use of small a and b can result in high posterior weights for $k = 1$ and $k = n$ (where k is the number of clusters, and n the number of individuals). Instead they propose an alternative method which results in $a = b = exp(-0.033n)$, giving a prior mean of unity with increasing standard deviation with larger n. The appeal of the method by Murugiah and Sweeting (2012) is that the prior gives less rigid adherence to G_0 with more data. In cases with small n it will generally be futile to search for, e.g. multimodality, so there is no

gain in allowing overly flexible realisations of G_0.

I combine work by Escobar and West (1995) and Murugiah and Sweeting (2012) to update the precision parameter: methods developed by Murugiah and Sweeting (2012) to update the hyper-parameters are incorporated into the sampling framework developed by Escobar and West (1995).

Escobar and West (1995) describe how α can be updated by incorporating an auxiliary variable η into the Gamma prior. The formula for updating α is expressed as a mixture of two gamma posteriors, with the conditional mixing parameter for α and k, a simple Beta distribution. They found that $p(\alpha|k)$ is the marginal distribution from a joint distribution for α and continuous quantity, η, such that

$$p(\alpha, \eta|k) \quad \propto \quad p(\alpha)\alpha^{k-1}(\alpha + n)\eta^{\alpha}(1-\eta)^{n-1}$$

where η is sampled from a Beta distribution: $(\eta|\alpha, k) \sim B(\alpha+1, n)$. Taking the conditional posteriors

$$\begin{aligned} p(\alpha|\eta, k) \quad &\propto \quad \alpha^{a+k-2}(\alpha + n)e^{-\alpha(b-log(\eta))} \\ &\propto \quad \alpha^{a+k-1}e^{-\alpha(b-log(\eta))} + n\alpha^{a+k-1}e^{-\alpha(b-log(\eta))} \end{aligned}$$

when $\alpha > 0$ this reduces to a mixture of two Gamma densities

$$(\alpha|\eta, k) \quad \sim \quad \pi_{\eta}G(a+k, b-log(\eta)) + (1-\pi_{\eta})G(a+k-1, b-log(\eta))$$

where $\pi = \frac{(a_{\alpha}+k-1)}{(n(b_{\alpha}-log(\eta))+a_{\alpha}+k-1)}$.

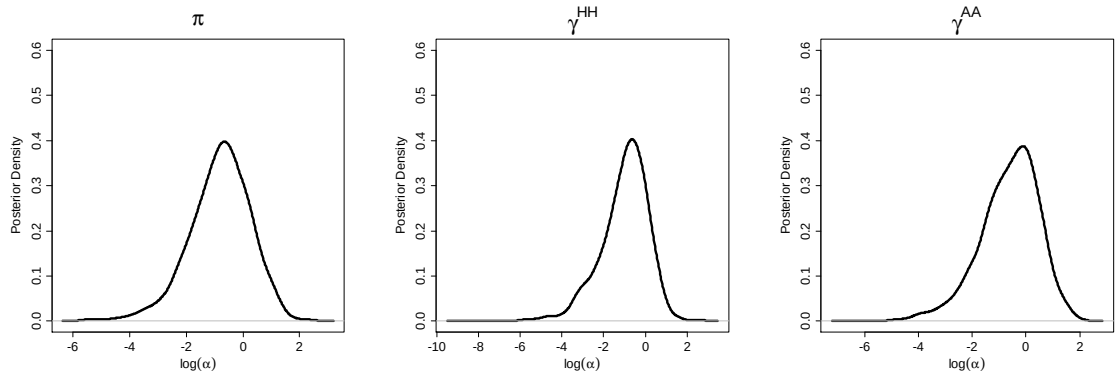

Figure 5.2: Smoothed density curve of posterior estimates for $\log(\alpha)$ for each parameter.

5.4 Simulation testing

The two-state model was used to test the Dirichlet process using a synthetic data set with 30 individuals, each with 1000 length capture history. We assumed individuals came from (randomly) one of two discrete groups: π= 0.82 or 0.96; γ^{HH}= 0.88 or 0.98; γ^{AA} = 0.8 or 0.95. Three separate chains were run for 15000 iterations. The chains were arbitrarily thinned to every 2nd update and combined to form one chain of 22500 posterior samples. The chains were thinned to reduce any autocorrelation between successive samples (Gilks et al., 1996).

The posterior density of $\log(\alpha)$ displays standard unimodal form (Figure 5.2). The posterior distribution of k (the number of clusters) indicated two clusters for each parameter. Figure 5.3 displays the posterior density for each of π , γ^{HH} and γ^{AA}, with posterior estimates for each parameter clustered around the two true values used for simulation.

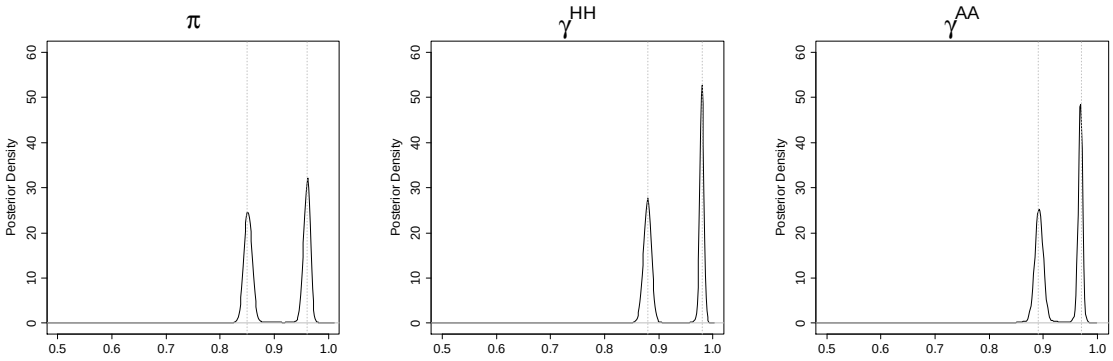

Figure 5.3: Results from 22500 updates combined from three independent chains. Grey dashed vertical lines indicate true value used in data simulation and posterior density of parameters appear to cluster around true values.

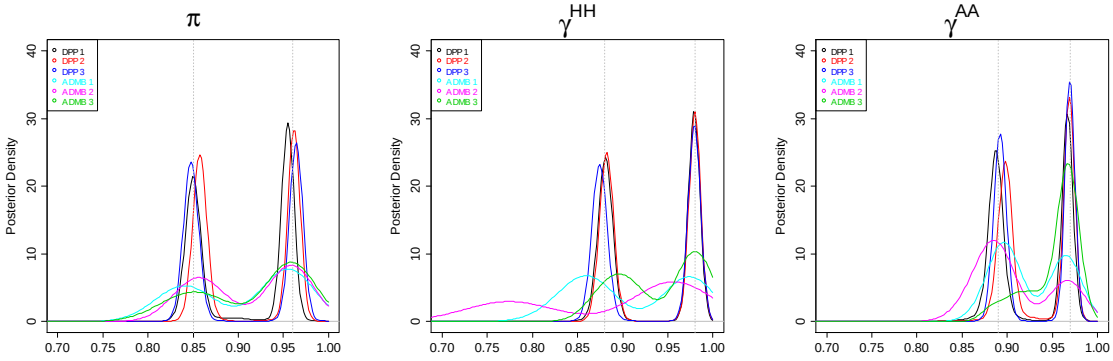

Figure 5.4: Results from three independent data sets estimated using posterior estimates from the Dirichlet process prior and individual posterior modes from ADMB-RE.

5.4.1 Comparison with ADMB

We compared the performance of this method against the Laplace approximation using three independent ADMB estimations (using the parameters specified in section 5.4). Figure 5.4 indicates that the Dirichlet process prior was more accurate in identifying the distinct clusters. Although ADMB managed to identify the bimodality in the posterior estimates, the certainty around these values was weaker and more variable between simulations.

5.4.2 Limits of Dirichlet process prior

The following example is intended to highlight the potential limits of the Dirichlet process in identifying clusters. Data was simulated for 30 animals with 1000 length capture history and run for 15000 iterations, with the first 5000 discarded due to burn-in. Three groups were assumed for both π ($\pi = 0.6, 0.85, 0.96$) and γ^{HH} ($\gamma^{HH} = 0.5, 0.8, 0.95$), and two groups for γ^{AA}($\gamma^{AA} = 0.89, 0.97$). Individuals were randomly assigned to a group for each parameter.

Figure 5.5 indicates the inability of the Dirichlet process prior to distinguish between low π and low γ^{HH}. The results indicate that the lowest true group in π ($p = 0.6$) could not be distinguished, and that the lowest estimated group in γ^{HH} was lower than the actual true values used in data simulation ($p = 0.5$). This result is unsurprising due to uninformative data and the resulting inability to distinguish between not being present and not being seen (see section 4.1 in chapter 3). At higher probabilities the posterior density of parameters appeared to cluster around the true discrete values used in data simulation.

5.4.3 Unimodal distributions

One concern with the use of Dirichlet process prior is the potential for spurious multimodality when in fact none is present. I used 10 independent synthetic data sets with 30 animals each with 1000 length capture history. For each parameter (π, γ^{HH} and γ^{AA}), synthetic data was simulated from a Normal distribution with low variance, $N(2, 0.1)$. The MCMC algorithm was run for 10000 iterations. There was no evidence of bi-modality in the results (Figure 5.6).

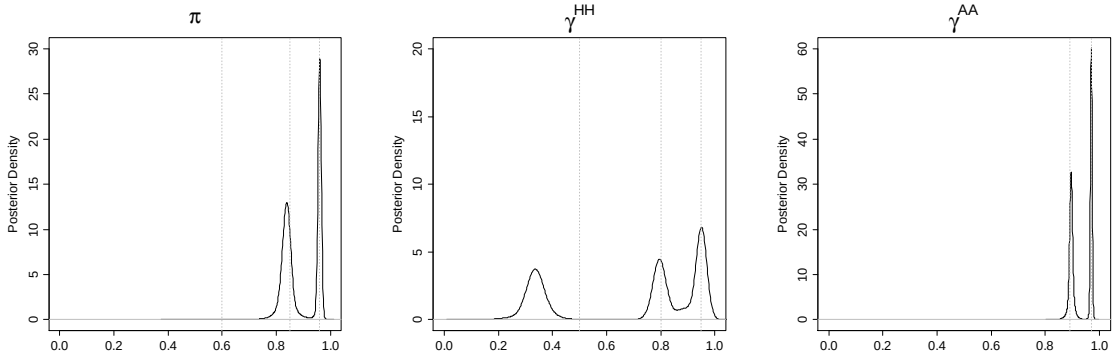

Figure 5.5: Results from 10000 iterations indicating the inability of Dirichlet process prior to distinguish low π and low γ^{HH}. Grey dashed vertical lines indicate the true values used in data simulation.

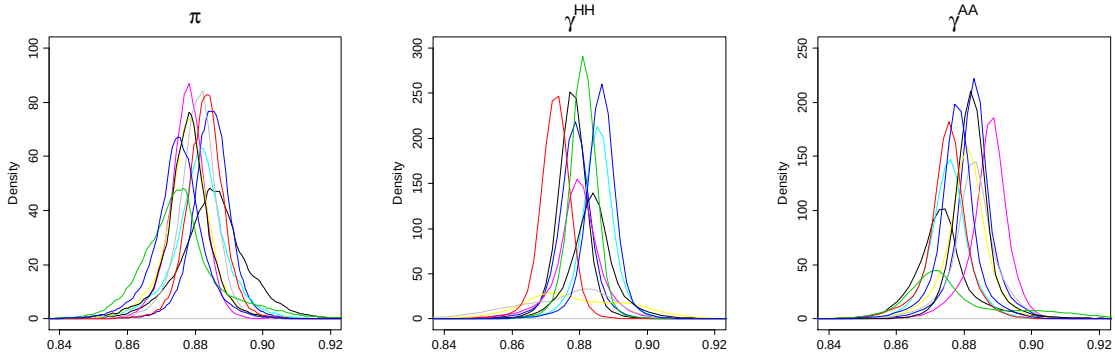

Figure 5.6: Results from 10 independent data sets to test ability of Dirichlet process prior to identify unimodal distribution. Individual parameter values generated using a logit-link and a Normal distribution with low variance.

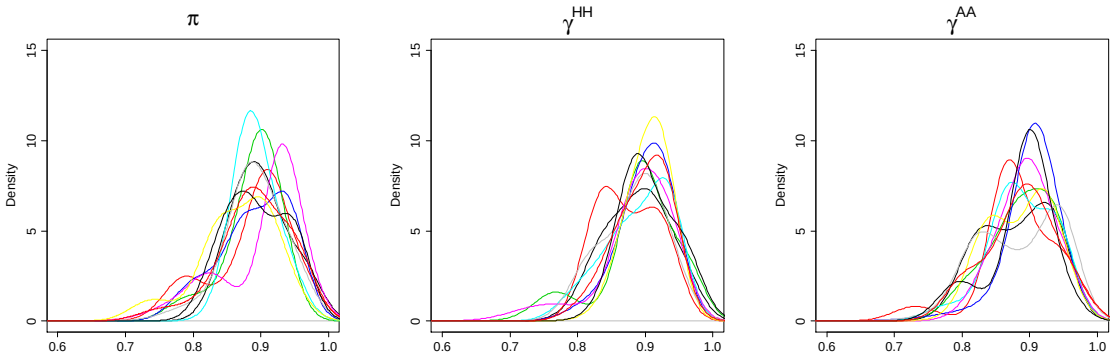

Figure 5.7: Results from 10 independent data sets to test ability of Dirichlet process prior to identify unimodal distribution. Individual parameter values generated from a Beta distribution.

Additional simulations from unimodal Beta distributions ($Beta(40, 5)$) indicate that Dirichlet process prior can sometimes produce some slight multimodality when simulated from unimodal distributions (Figure 5.7). However, on exploration (not shown) it was apparent that the posterior estimates generally reflected the distribution of true values used in simulation. That is, with small sample size from a Beta distribution there were some 'outlying' individual-level probabilities which were captured by the posterior estimates of the Beta-Binomial model. Correlations, between mean individual posterior estimates and true values used in simulation, were high for each parameter. Mean correlations for the 10 simulations were: $\pi = 0.89$, $\gamma^{HH} = 0.91$ and $\gamma^{AA} = 0.94$.

5.5 Application to North Atlantic humpback whales

5.5.1 Model

The full three-state model, including death (the same structure as used in chapter 3 and 4), was applied to a sub-set of 176 North Atlantic hump-

back whales sighted more than 30 times in the Stellwagen National Marine Sanctuary (SBNMS) between 1979 and 2005. A subset of the real data was used, rather than all 1147 whales as per chapter 3, due to the computational speed of the MCMC sampler. The three-state model uses same hierarchical implementation of the model outlined in chapter 4, with the difference that the individual parameter values are updated using the Dirichlet process prior (described above). For clarity several sections from chapter 3 and 4 are briefly reiterated here.

With three states the nine elements of the transition matrix can be written in terms of just three parameters γ^{HH}, γ^{AA}, and γ^{D} (respectively the probabilities of staying Here, staying Away and Dying) as follows:

$$\gamma = \begin{pmatrix} \gamma^{HH}\left(1-\gamma^{D}\right) & \left(1-\gamma^{HH}\right)\left(1-\gamma^{D}\right) & \gamma^{D} \\ \left(1-\gamma^{AA}\right)\left(1-\gamma^{D}\right) & \gamma^{AA}\left(1-\gamma^{D}\right) & \gamma^{D} \\ 0 & 0 & 1 \end{pmatrix}$$

where it is assumed that the probability of death (which is very low relative to the other transition rates) does not depend on whether the animal is Here or Away.

Only data from the first sighting of each whale onwards are used. As sighting effort is focused in the middle of the year, we included all sightings from the 18th week of the year through to the 43rd week. The probability of survival, P_{surv}, over the remaining 26 week period was calculated as $P_{surv} = (1-\gamma^{D})^{26}$. An extra parameter q was introduced for the chance of being present in the SBNMS at the start of the season. We calculated the

probability of each state in the first week of the new year to be:

$$\mathbb{P}(S_t|S_{t-1}) = \begin{pmatrix} q * P_{surv} & (1-q) * P_{surv} & 1 - P_{surv} \\ q * P_{surv} & (1-q) * P_{surv} & 1 - P_{surv} \\ 0 & 0 & 1 \end{pmatrix} * \mathbb{P}(S_{t-1})$$

where $\mathbb{P}(S_{t-1})$ is the vector of state probabilities in the last week of the previous year.

Individual-level random effects were included on each of π, γ^{HH} and γ^{AA} and are updated exactly as described for the two-state model applied to synthetic data (see section 5.2.1). We assume individual-level parameters to be consistent over time but have allowed for population-level annual variation in probability of remaining Here using logit-links: $\text{logit}\gamma_{i,yr}^{HH} = \beta_{yr} + \gamma_i^{HH}$. Updates to β_{yr}, γ^D (death) and q are assumed to be fixed (not individually variable) and updated using the same structure as described in chapter 4 (reiterated below for clarity).

5.5.2 Updates to population-level fixed effects

As outlined in chapter 4 we have incorporated results from ADMB to facilitate updates to each of γ^D, q and β_{yr}. The structure for update is the same as chapter 4.

5.5.3 Results

One chain was run for 100000 iterations with the first 10000 discarded to burn-in. Density plots of $\log(\alpha)$ indicate expected unimodal density (Figure 5.8). The posterior distribution of the number of clusters indicated more variation for π compared to both γ^{HH} and γ^{AA}. Figure 5.9 indicates some

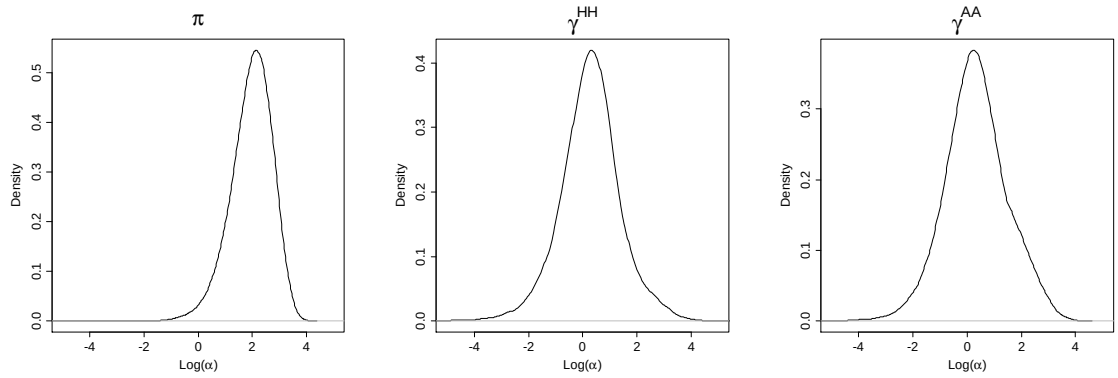

Figure 5.8: Density plot for $\log(\alpha)$, the precision parameter in Dirichlet process prior

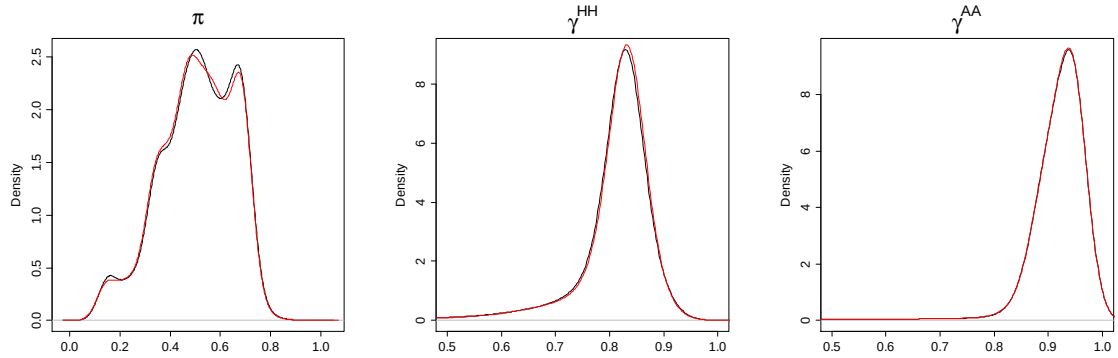

Figure 5.9: Density for π, γ^{HH} and γ^{AA} for 90000 iterations for 176 whales seen more than 30 times between 1979 and 2005. Black density curves indicate first 45000 iterations of the MCMC chain, and red the second 45000 iterations.

multimodality in both halves of the MCMC chain for π; less variation was evident in both state probabilities compared to the probability of observation.

5.5.4 Comparison: ADMB, Beta-Binomial and Dirichlet process

Comparison of the three methods indicate similar results for individual-level random effects for γ^{AA}. The functional form of the random effects for π is

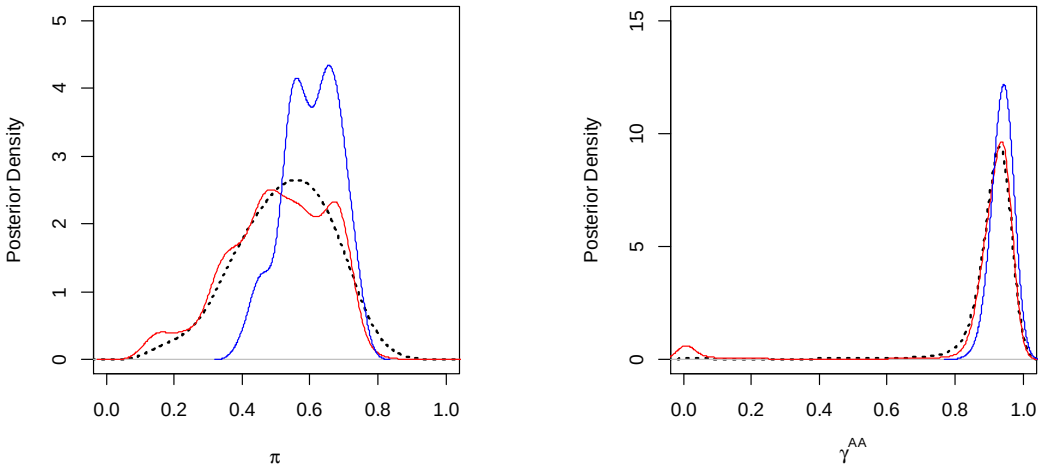

Figure 5.10: Comparison of individual posterior estimates for π and γ^{AA} for ADMB (blue line), Beta-Binomial (black dashed line) and Dirichlet process prior (red line).

different between the three methods: the Dirichlet process prior indicates some multimodality in π, which was less evident in both ADMB and the Beta-Binomial method (Figure 5.10). Results were also similar for posterior estimates of γ^{HH} and β_{yr} (Figure 5.11).

5.6 Discussion

In some studies, covariates may adequately explain the majority of individual heterogeneity present in the data. However, in some cases and for certain species (notably cetaceans), it is unrealistic to expect to be able to collect all necessary covariates, or in fact to know which covariates explain the heterogeneity. Accurately capturing this latent heterogeneity is important to ensure accurate analysis of mark-recapture data. The use of the hidden Markov model combined with the Dirichlet process prior, provides a

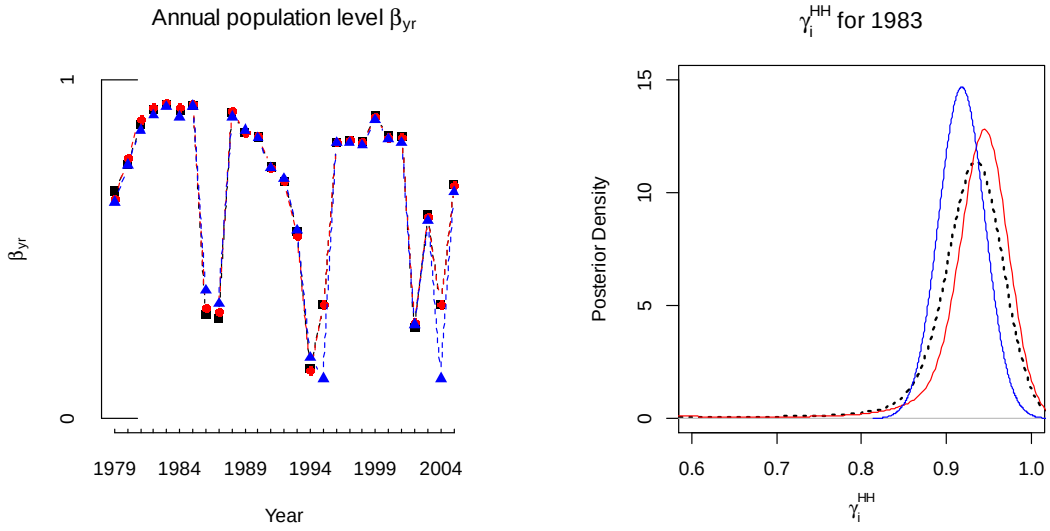

Figure 5.11: Comparison of posterior estimates for β_{yr}, andγ_i^{HH} in year 1983 for: ADMB (blue triangle/line), Beta-Binomial (black squares/line) and Dirichlet process prior (red circles/line).

powerful tool for capturing latent individual heterogeneity.

Our results show the Dirichlet process prior was able to accurately capture multimodality in three measures of individual heterogeneity: probability of observation, probability of remaining in the marine sanctuary and probability of remaining away. Through simulation studies we were able to explore the accuracy, and limits, of the Dirichlet process prior to distinguish multiple groups using this framework. We found in certain areas of parameter space, the Dirichlet process prior was capable of capturing up to three distinct groups. However, the model was not capable of distinguishing between low probability of observation and low probability of remaining in the marine park. This aliasing is unsurprising (see chapter 3) and is due to the lack of information contained in the capture histories, rather than due to the Dirichlet process prior.

Using synthetic data we compared the Dirichlet process prior with ADMB-

RE. Our results indicate that ADMB-RE was able to capture the bi-modality in the synthetic data. However, compared to the Dirichlet process model, the density around the true values was less and the posterior distribution of the estimates wider.

In application to North Atlantic humpback whales, we used data for a subset of whales frequently observed during the years 1979 to 2005. Posterior estimates of the state transition probabilities were less varied than for the probability of observation. There was some multimodality apparent in the probability of observation which could have implications for abundance estimates in this population. The variation in both the state transition probabilities, although small in application to the subset of data, would result in substantial differences in proportion of time spent in the marine park. This has implications in the ability to predict the long term usage of the marine park and for population survival and growth.

If we are primarily interested in the estimation of individual estimates, rather than population-level estimates, then the complexity of the Dirichlet process is required. There are several extensions and applications of the Dirichlet process which were not explored here but are important considerations and interesting areas for future exploration.

In comparison to parametric distributions, the Dirichlet process allows for multiple modes in both the observation and state process. Heterogeneity in detection in mark-recapture data has been a hurdle in the accurate estimation of abundance. With the potential to identify multiple modes in the probability of observation, the Dirichlet process has the potential to give more accurate estimates of abundance. The Dirichlet process also has important application to more effective marine spatial planning. Results in chapter 2 indicated the population consequences of individual heterogeneity

in the context of marine spatial management. In chapter 3, the stationary distribution of the transition matrix was used to estimate the individual proportion of time spent in the marine reserve. The Dirichlet process provides a method to more accurately capture the individual behaviour, which translates into more accurate estimations of proportion of time spent in the marine reserve.

Additional interesting applications could involve exploration of correlations between individual random effects: for example, multiple behavioural modes might indicate that if an individual is often away, they may also be more likely to be infrequently observed. With the addition of random effects onto arrival time each year it would be interesting to see the correlation between arrival and departure, and arrival and length of stay in the marine park. It would be worthwhile to investigate the ability of the Dirichlet process to model this, or other, correlations in behaviour.

The development of Bayesian hierarchical models has been the focus of much effort in mark-recapture research (King, 2012). Despite this, non-parametric approaches have received little attention. We have presented a hierarchical hidden Markov model which allows for both process and observation error and have incorporated the Dirichlet process prior to account for individual heterogeneity on both the observation and process components. We anticipate that this powerful addition to mark-recapture analysis will be useful in application to other problems by allowing for accurate estimation of multiple behavioural modes.

Chapter 6

Conclusion

Application to marine protected areas

For many mark-recapture studies, the main motivation for estimating individual heterogeneity is to reduce bias in estimates of abundance and survival. To date there has been little consideration of the ramifications of individual heterogeneity in application to other areas, such as marine spatial management. For this reason we chose to focus our initial exploration on the population consequences of individual heterogeneity in the context of a marine protected area (MPA) (see chapter 2). If a MPA is too large it could be economically infeasible to manage properly. Conversely, too small an area will be ineffective at providing adequate protection. Determining the most effective area to place the MPA boundaries, or the effectiveness of an existing MPA, depends in part on individual spatial use. Hence, in chapter 2, using a basic population dynamics model and differential survival inside and outside the MPA, I explored the population consequences of individual heterogeneity in spatial usage. The results highlighted the importance of accounting for individual heterogeneity when assessing the effectiveness of MPAs. While MPAs are often designed primarily with the aim of conserving

animals within the MPA itself, they can also be of benefit in maintaining numbers outside.

The approach of using the Markovian movement model to describe individual-level propensity to use a protected area has applications throughout marine and terrestrial ecology. Determining the degree to which individuals vary in their use of an area should enhance the ability to assess the effectiveness of a MPA. Using this framework of a movement model, chapters 3, 4, and 5 focused on the development of methods using random effects, to estimate individual heterogeneity in observation and also in probability of remaining in or out of the marine reserve.

Random effects

In all mark-recapture studies, some animals have longer capture histories than others. One appeal of using random effects to model individual heterogeneity, lies in the ability for animals with little data to borrow strength from those with lots of data. Other limitations can be seen with the approach of using per individual fixed effects, not least that it is difficult to discuss population-level behaviour of certain parameters. Typically in mark-recapture studies, a high proportion of the population are unobserved; no statistical inference can be made about the parameter values of the unseen animals if a fixed effect approach is used. Assuming a continuous distribution of individual-level random effects is a more natural, but more complex approach, as it introduces the complexity of dealing with these random effects. This has been a major impediment in the mark-recapture field, amongst others.

A variety of methods and software have been used to estimate the individual-level random effects in mark-recapture data (Gimenez and Choquet, 2010;

Pledger et al., 2003; Royle and Dorazio, 2008). This thesis has considered three methods. In chapter 3 I presented a multi-state model where the unobservable states follow a hidden Markov model, which allows for individual heterogeneity in both detection probability and site fidelity. This framework is used throughout the thesis. The model was developed in ADMB-RE, an open source software package. The appeal of ADMB-RE lies in its speed and ease of implementation for the user: the combination of automatic differentiation and Laplace approximation results in a powerful tool which avoids complex calculations and computations on the user's part. ADMB-RE has had little application to mark-recapture data, but it provides an excellent addition to the suite of programs already available and being widely used (i.e WinBUGs, E-SURGE, Program MARK, and R). Given the relative ease, flexibility, and use of ADMB, and the large project support available, I anticipate increased use of this invaluable software in mark-recapture research.

Another obvious approach for building complex random effect models for individual heterogeneity is Markov chain Monte Carlo (MCMC). In chapter 4 I developed a MCMC sampler which incorporates individual heterogeneity using parametric random effects. I extended this in the fifth chapter by using the non-parametric Dirichlet process prior. My aim was to develop several methods, and compare their ease and implementation. The Dirichlet process prior was considered the ultimate goal as it is both a flexible extension to parametric methods and can also be thought of as the infinite limit of the discrete groups approach. Despite its appeal, the Dirichlet process has had little application in mark-recapture research (one good counter example is the paper by Dorazio et al., 2008). The incorporation of the Dirichlet process prior into the hidden Markov model framework results in a powerful tool for mark-recapture data, allowing both process and observation noise and the

estimation of individual variation.

In chapter 5, the Laplace approximation and the Dirichlet process prior were compared in their ability to capture multimodality in synthetic data sets and also in application to the real data. The comparison of ADMB to the Dirichlet process, indicated ADMB was able to accurately identify multimodality, but with more variance in posterior estimates than the Dirichlet process. This raises the question of which approach should be used: the relatively more easily implemented Laplace approximation using ADMB or the more complicated, yet more accurate, Dirichlet process. The Dirichlet process is clearly a more sensitive tool, but it should be noted that our implementation of both the Beta-Binomial and the Dirichlet process model required results from ADMB. Implementation of ADMB may potentially miss some aspects of heterogeneity which would be captured under the Dirichlet process, but the implementation is quicker, and extension of the model is more straightforward than the MCMC approach. If individual estimates are the main interest for a study, then the Dirichlet process is the most appropriate of these methods as the detail of different behaviours would be lost in the first two methods. However, if we are most interested in the population level variation then ADMB would be an adequate approach, and certainly the most straight forward.

When contemplating fitting a variety of models, there are always costs or trade-offs to consider: fitting a model that is too simple may result in increased bias or variance, whereas fitting a model that is too complex may result in high prediction error or computational difficulties. In this context, one aspect of complexity is specifying how many model parameters are allowed to have latent heterogeneity. In chapter 3, I explored the costs of not accounting for heterogeneity when in fact it was present, and vice versa al-

lowing for heterogeneity in the model when there was none in the data. The results indicated that there was little cost associated with having a model which was more complex than the data at hand, as random effects were correctly estimated with low variance when not present in the data. These models perform well when both γ^{HH} and γ^{AA} are both transition parameters are more than 0.5. During model exploration simulations indicated that the model parameters γ^{HH} and γ^{AA} may not be identifiable when both are below 0.5 and the estimation process performs poorly for parameters values in this parameter space. The capture histories which result from low transition probabilities, and thus high transition rates, makes it difficult to accurately estimate proportion of time spent in the reserve as it is difficult to distinguish between non-detection and being Away.

This thesis has successfully estimated several measures of individual heterogeneity for over 1000 individuals (see chapter 3). One clear advantage of ADMB over MCMC during model testing, is the ability to run multiple parameter estimations from simulated data; a time consuming and computer intensive process with MCMC. Application of the methods developed in this thesis in WinBUGs would appear to be possible (both HMMs and Dirichlet processes have been implemented separately), and the code is arguably more intuitive than ADMB code, but a lack of flexibility and speed makes it an impractical choice for simulation testing as in chapter 3. In regards to the MCMC framework developed in chapters 4 and 5, the model was coded using R and C++. This was an intentional choice, despite the availability of the DPpackage in R, in order to allow flexibility in application and extension of the models. It is unclear whether the DPpackage would integrate with the process model coded in C++, and it does not allow for updates to the precision parameter (as per Murugiah and Sweeting, 2012) as implemented

in this thesis.

Future extensions and research

As the focus of this thesis has been methodological developments, there has been relatively little focus or incorporation of biological covariates (e.g. sex, age, length) or model selection and checking. Throughout this thesis I generally neglected any direct comparison of models. This was largely due to the focus on the technical development of the methods and the incorporation of the random effects.

The Empirical Bayes approach in ADMB results in an approximate marginal likelihood, with random effects integrated out. This likelihood can then be used for standard model selection techniques such as AIC. This approach is more difficult in the fully hierarchical Bayesian approach (Hodges and Sargent, 2001). Model selection criteria such as BIC and DIC are well known to be problematic for complicated hierarchical models, hidden Markov models and models with parameters with non-normal posterior distributions (Scott, 2002; Shirley et al., 2010; Spiegelhalter et al., 2002).

Model selection and checking in the hidden Markov model literature has focused mainly on the selection of optimal number of hidden states. In less complicated models, with only fixed effects, several studies have explored various methods for model checking and selection (see Jonsen et al., 2012; Scott, 2002; Zucchini and MacDonald, 2009). The inclusion of individual-level random effects into the models greatly complicates the process and this is definitely an area for future research.

There was also relatively little focus or incorporation of biological covariates into the models. In most studies, there will be collection of data

on individual-level covariates which could go a long way towards explaining the individual differences present in the population. However, in most cases, there will still be some latent individual heterogeneity, due to the inherent differences between individuals. Random effects should be used in combination with available covariates in order to accurately estimate true latent individual variation.

Along with incorporation of covariates, several extensions to the model developed in this thesis would improve the biological realism of the process model. Exploration of the North Atlantic humpback whale data indicated that the whales displayed evidence of variation in yearly arrival times and patterns. Several extensions could be included in application to this data, for example: an individual-level random effect that allows for differences in arrival times (i.e. a whale may be consistently late to arrive in the SBNMS each feeding season); and annual population-level variation in γ^{HH} could be modeled as a random effect (it is currently a fixed effect). Inclusion of these extensions would improve the model, but introduce the additional complexity of crossed random effects. Crossed random effects would mean that in both ADMB and MCMC the separable nature of the problem could no longer be exploited.

One major advantages of modeling via a hidden Markov model, is that it allows for unobservable states (for example breeding or physiological state). Although the focus in this thesis was on geographical location (in/out of the marine park), the methods developed in this thesis can be extended to incorporate other states such as breeding, age, or body weight. Such extension would allow the estimation of an alternative form of individual heterogeneity in terms of fitness or quality. Various measures of individual fitness include timing of reproduction, reproductive effort, maternal investment, longevity,

or paternal ability condition (Moyes et al., 2009).

Although applied only to mark-recapture data, the methods developed for estimating individual-level heterogeneity here are also applicable to other forms of data being widely used within ecology, e.g. individual animal movement recorded using telemetry or GPS (Jonsen et al., 2012; Sharples et al., 2012; Block et al., 2005). These techniques also allow for a more detailed understanding of individual behaviour. The methods developed in this thesis to estimate individual heterogeneity could be a useful addition to the methods already being used in this area (e.g. Langrock et al. 2012 who incorporate some random effects, and Patterson et al., 2008). For example, in application to understanding population structure of harbour seals, the Dirichlet process could be useful in identifying different foraging behaviour where separate colonies exist (e.g. Sharples et al., 2012).

This thesis has highlighted the importance of accounting for individual heterogeneity and successfully applied novel statistical methods which facilitate its estimation. The application of the Dirichlet process avoids the awkward model selection of the discrete groups approach, and the inflexibility of imposing a prior parametric form associated with other methods. Individual heterogeneity is known to matter for the accurate estimation of classical mark-recapture quantities such as abundance and survival. But its importance does not end there, we show, for example, how heterogeneity in spatial use can affect the performance of a MPA. The methods have the potential to aid in design, assessment, and monitoring of MPAs for individually identifiable animals. The extension to incorporate a non-parametric prior unifies two statistical modeling approaches and allows for unimodal or multimodal behaviour in several measures of individual-level heterogeneity. Although there have been significant computational challenges in getting the

methods working, we anticipate that ADMB-RE and the Dirichlet process prior will see more and more use in individual-based mark-recapture analysis and throughout ecology more broadly.

Bibliography

Arnason, A. N. 1972. Parameter estimates from mark-recapture experiments on two populations subject to migration and death. Researches in Population Ecology **13(2)**:97–113.

Arnason, A. N. 1973. The estimation of population size, migration rates, and survival in a stratified population. Researches on Population Ecology **15(1)**:1–8.

Ashe, E., D. Noren, and R. Williams. 2010. Animal behaviour and marine protected areas: incorporating behavioural data into the selection of marine protected areas for an endangered killer whale population. Animal Conservation **13**:196–203.

Barco, S., W. McLellan, J. Allen, R. Asmutis-Silva, R. Mallon-Day, E. Meagher, D. Pabst, J. Robbins, R. Seton, W. Swingle, M. Weinrich, and P. Clapham. 2002. Population identity of humpback whales (megaptera novaeangliae) in the waters of the us mid-atlantic states. Journal of Cetacean Research and Management **4**:135–141.

Barry, S. C., S. P. Brooks, E. Catchpole, and B. Morgan. 2003. The analysis of ring-recovery data using random effects. Biometrics **59(1)**:54–65.

Bayes, T. 1763. An essay towards solving a problem in the doctrine of chances. Philosophical Transactions of the Royal Society **53**:370–418.

Beausoleil, N., D. Mellor, and K. Stafford, 2004. Methods for marking new zealand wildlife: Amphibians, reptiles and marine mammals. Technical report, Department of Conservation *Te Papa Atawhai*.

Blackwell, D. and J. MacQueen. 1973. Ferguson distributions via polya urn shcheme. Annals of Statistics **1**:353–355.

Block, B., S. Teo, A. Walli, A. Boustany, C. Farwell, H. Dewar, K. Weng, and T. Williams. 2005. Electronic tagging and population structure of atlantic bluefin tuna. Nature **434**:1121–1127.

Bolnick, D., R. Svanbäck, J. Fordyce, L. Yang, J. Davis, C. Hulsey, and M. Forister. 2003. The ecology of individuals: incidence and implications of individual specialization. American Naturalist **161**:1–28.

Breslow, N. and D. Clayton. 1993. Approximate inference in generalized linear mixed models. Journal of the American Statistical Association **88(421)**:9–25.

Burnham, K. P. and W. S. Overton. 1978. Estimation of size of a closed population when capture probabilities vary among animals. Biometrika **65(3)**:625–633.

Clapham, P., L. Baraff, C. Carlson, M. Christian, D. Mattila, C. Mayo, M. Murphy, and S. Pittman. 1993. Seasonal occurrence and annual return of humpback whales, *Megaptera novaeangliae*, in the sourthern gulf of maine. Canadian Journal of Zoology **71**:440–443.

Clapham, P. and C. Mayo. 1987. Reproduction and recruitment of individually identified humpback whales, *Megaptera novaeangliae*, observed in massachusetts bay, 1979-1985. Can. J. Zool. **65**:2853–2863.

Clapham, P. and J. Mead. 1999. *Megaptera novaeangliae*. Mammalian Species **604**:1–9.

Conn, P. B. and E. G. Cooch. 2009. Multistate capture-recapture analysis under imperfect state observation: an application to disease models. Journal of Applied Ecology **46(2)**:486–492.

Cubaynes, S., C. Lavergne, E. Marboutin, and O. Gimenez. 2012. Assessing individual heterogeneity using model selection criteria: how many mixture components in capture-recapture models? Methods in Ecology and Evolution **3**:564–573.

D'Agrosa, C., C. Lennert-Cody, and O. Vidal. 2000. Vaquita bycatch in mexico's artisanal gillnet fisheries: Driving a small population to extinction. Conservation Biology **14(4)**:1110–1119.

Dorazio, R. M. 2009. On selecting a prior for the precision parameter of dirichlet process mixture models. Journal of Statistical Planning and Inference **10**:10–16.

Dorazio, R. M., B. Mukherjee, L. Zhang, M. Ghosh, H. Jelks, and F. Jordan. 2008. Modeling unobserved sources of heterogeneity in animal abundance using a dirichlet process prior. Biometrics **64(2)**:635–644.

Escobar, M. and M. West. 1995. Bayesian density estimation and inference using mixtures. American Statistical Association **90**:577–588.

Fearnhead, P. 2004. Particle fileters for mixture models with an unknown number of components. Statistics and Computing **14**:11–21.

Ferguson, T. 1973. A bayesian analysis of some nonparametric problems. Annals of Statistics **1**:209–230.

Fournier, D. A., H. Skaug, J. Ancheta, J. Ianelli, A. Magnusson, M. Maunder, A. Nielsen, and J. Sibert. 2012. AD model builder: using automatic differentiation for statistical inference of highly parameterized complex nonlinear models. Optimization Methods & Software **27:2**:233–249.

Gelfand, A. 2000. Gibbs sampling. Journal of the American Statistical Association **95**:1300–1304.

Gelfand, A. and A. Smith. 1990. Sampling-based approaches to calculating marginal densities. Journal of the American Statistical Association **85**:398–409.

Gell, F. and C. Roberts. 2003. Benefits beyond boundaries: the fishery effects of marine reserves. Trends in Ecology and Evolution **18(9)**:448–455.

Gerrodette, T. and L. Rojas-Bracho. 2011. Estimating the success of protected areas for the vauqita, *Phocoena sinus)*. Marine Mammal Science **27**:E97–E100.

Gilks, W., S. Richardson, and D. Spiegelhalter. 1996. Markov Chain Monte Carlo in Practice. Chapman and Hall /CRC.

Gimenez, O. and R. Choquet. 2010. Individual heterogeneity in studies on marked animals using numerical integration: capture-recapture mixed models. Ecology **91(4)**:951–957.

Gimenez, O., J. D. Lebreton, J.-M. Gaillard, R. Choquet, and R. Pradel. 2012. Estimating demographic parameters using hidden process dynamic models. Theoretical Population Biology **In Press**.

Gormley, A., E. Slooten, S. Dawson, R. Barker, W. Rayment, S. du Fresne, and S. Brager. 2012. First evidence that marine protected areas can work for marine mammals. Journal of Applied Ecology **49**:474–480.

Hammond, P. S. 1986. Estimating the size of naturally marked whale populations using capture-recapture techniques. Reports of the International Whaling Commission. **(Special Issue) 8**:253–282.

Hammond, P. S. 1990. Heterogeneity in the gulf of maine? estimating humpback whale population size when capture probabilities are not equal. Reports of the International Whaling Commission **(Special Issue) 12**:135–139.

Hastings, W. 1970. Monte carlo sampling methods using markov chains and their applications. Biometrika **57**:97–109.

Hodges, J. and D. Sargent. 2001. Counting degrees of freedom in hierarchical and other richly parameterized models. Biometrika **88**:367–379.

Hoelzel, A. 1998. Genetic structure of cetacean populations in sympatry, parapatry, and mixed assemblages: Implications for conservation policy. Theamerican Genetic Association **89**:451–458.

Hooker, S., A. Canadas, K. Hyrenbach, C. Corrigan, J. Polovina, and R. Reeves. 2011. Making protected area networks effective for marine top predators. Endangered Species Research **13**:203–218.

Hoyt, E. 2011. Marine Protected Areas for Whales, Dolphins and Porpoises: A World Handbook for Cetacean Habitat Conservation and Planning. Earthscan, London and Washington.

Huggins, R. M. and P. S. F. Yip. 2001. Note on nonparametric inference for

capture-recapture experiments with heterogeneous capture probabilities. Statistica Sinica **11(3)**:843–853.

Ishwaran, H. and L. James. 2001. Gibbs sampling methods for stick-breaking priors. Journal of the American Statistical Association **96**:161–173.

IUCN, 1994. Guidelines for protected area management categories. Technical report, CNPPA with the assistance of WCMC, IUCN.

Jain, S. and R. M. Neal, 2004. A split-merge markov chain monte carlo procedure for the dirichlet process mixture model. Technical report, Technical report, Department of Statistics, University of Toronto.

Johnson, A., G. Salvador, J. Kenney, J. Robbins, S. Kraus, S. Landry, and P. Clapham. 2005. Fishing gear involved in entanglements of right and humpback whales. Marine Mammal Science **21(4)**:635–645.

Jonsen, I., M. Basson, S. Bestley, M. Bravington, T. A. Patterson, M. Pedersen, R. Thomson, U. Thygesen, and S. Wotherspoon. 2012. State-space models for bio-loggers: A methodological road map. Deep-Sea Research **In Press DOI:10.1016/j.dsr2.2012.07.008**.

Kaplan, D., S. Planes, C. Fauvelot, T. Brochier, C. Lett, N. Bodin, F. Le Loc'h, Y. Tremblay, and J.-Y. Georges. 2010. New tools for the spatial management of living marine resources. Current Opinion in Environmental Sustainability **2**:88–93.

Katona, S. and J. Beard. 1990. Populatoin size, migrations, and feeding aggregations of the humpback whale *Megaptera novaeangliae* i nthe western north atlantic ocean. Reports of the International Whaling Commission. **(Special Issue) 12**:295–306.

Katona, S. K. and H. P. Whitehead. 1981. Identifying humpback whales using their natural markings. Polar Record **20(128)**:439–441.

Keith, J., D. Kroese, and G. Y. Sofronov. 2008. Adaptive independence samplers. Statistics and Computing **18**:409–420.

Kerwath, S., E. Thorstad, T. Næsje, P. Cowley, F. Økland, C. Wilke, and C. Attwood. 2008. Crossing invisible boundaries: the effectiveness of the langebaan lagoon marine protected area as a harvest refuge for a migratory fish species in south africa. Conservation Biology **23**:653–661.

King, R. 2012. A review of bayesian state-space modelling of capture-recapture-recovery data. Interface Focus **2**:190.

Knowlton, A. and S. Kraus. 2001. Mortality and serious injury of northern right whales *(Eubalaena glacialis)* in the western north atlantic ocean. Journal of Cetacean Research and Management **(Special Issue) 2**:193–208.

Kyung, M., J. Gill, and G. Casella. 2010. Estimation in dirichlet random effects models. Annals of Statistics **38**:979–1009.

Laist, D., A. Knowlton, J. Mead, A. Collet, and M. Podesta. 2001. Collisions between ships and whales. Marine Mammal Science **17**:35–75.

Langrock, R., R. King, J. Matthiopoulos, L. Thomas, D. Fortin, and J. Morales. 2012. Flexible and practical modeling of animal telemetry data: hidden markov models and extensions. Ecology **93(11)**:2336–2342.

Lebreton, J. D., J. D. Nichols, R. Barker, R. Pradel, and J. Spendelow. 2009. Modeling individual animal histories with multistate capture-recapture models. In Hal Caswell, editor: Advances in Ecological Research **41**:87–173.

Lebreton, J. D. and R. Pradel. 2002. Multistate recapture models: modelling incomplete individual histories. Journal of Applied Statistics **29(1-4)**:353–369.

Liang, P., S. Petrov, M. Jordan, and D. Klein, 2007. A permutation-augmented sampler for dirichlet process mixutre models. *in* In *Proceedings of the International Conference on Machine Learning.*

Liu, Q. and D. Pierce. 1994. A note on guass-hermite quadrature. Biometrika **81(3)**:624–629.

MacEachern, S. and P. Muller. 1998. Estimating mixture of dirichlet process models. Journal of Computational and Graphical Statistics **7**:223–238.

Marsh, H. 2000. Evaluating management intitiatives aimed at reducing the mortality of dugongs in gill and mesh nets in the great barrier reef world heritage area. Marine Mammal Science **16(3)**:684–694.

Maunder, M., H. Skaug, D. Fournier, and S. Hoyle, 2008. Comparison of estimators for mark-recapture models: random effects, hierarchical bayes, and ad model builder. Pages 917–948 *in* In: Modeling Demographic Processes in Marked Populations. Eds. Thomson, D.L., Cooch, E.G., and Conroy, M.J. Environmental and Ecological Statistics, volume 3.

McNamara, J. and A. Houston. 1996. State-dependent life histories. Nature **380**:215–220.

Metropolis, N., A. W. Rosenbluth, M. N. Rosenbluth, and A. H. Teller. 1953. Equations of state calculations by fast computing machines. Journal of Chemical Physics **21**:1087–1092.

Moffitt, E., L. Botsford, D. Kaplan, and M. O'Farrell. 2009. Marine reserve networks for species that move within a home range. Ecological Applications **19(7)**:1835–1847.

Moffitt, E., J. White, and L. Botsford. 2011. The utility and limitations of size and spacing guidelines for designing marine protected area (mpa) networks. Biological Conservation **144**:306–318.

Moyes, K., B. Morgan, A. Morris, S. Morris, T. Clutton-Brock, and T. Coulson. 2009. Exploring individual quality in a wild population of red deer. Journal of Animal Ecology **78**:406–413.

Murugiah, S. and T. Sweeting. 2012. Selecting the precision parameter prior in dirichlet process mixture models. Journal of Statistical Planning and Inference **142**:1947–1959.

Navarro, D. J., T. L. Griffiths, M. Steyvers, and M. D. Lee. 2006. Modeling individual differences using dirichlet processes. Journal of Mathematical Psychology **50**:101–122.

Neal, R. M. 2000. Markov chain sampling methods for dirichlet process mixture models. Journal of Computational and Graphical Statistics **9**:249–265.

Norris, J. and K. Pollock. 1996. Nonparametric mle under two closed capture-recapture models with heterogeneity. Biometrics **52**:639–649.

Patterson, T. A., M. Basson, M. Bravington, and J. Gunn. 2009. Classifying movement behaviour in relation to environmental conditions using hidden markov models. Journal of Animal Ecology **78(6)**:1113–1123.

Patterson, T. A., L. Thomas, C. Wilcox, O. Ovaskainen, and J. Matthiopoulos. 2008. State-space models of individual animal movement. Trends in Ecology & Evolution **23(2)**:87–94.

Payne, P., D. Wiley, S. Young, S. Pittman, P. Clapham, and J. Jossi. 1990. Recent fluctuations in the abundance of baleen whales in the southern gulf of maine in relation to changes in prey abundance. Fish. Bull. **84**:41–47.

Pitman, J. 2006. Cobinatorial Stochastic Processes. Berlin: Springer-Verlag.

Pledger, S. 2000. Unified maximum likelihood estimates for closed capture-recapture models using mixtures. Biometrics **56**:434–442.

Pledger, S. and P. Phillpot. 2008. Using mixtures to model heterogeneity in ecological capture-recapture studies. Biometrical Journal **50(6)**:1022–1034.

Pledger, S., K. H. Pollock, and J. Norris. 2003. Open capture-recapture models with heterogeneity: I. cormack-jolly-seber model. Biometrics **59(4)**:786–794.

Pradel, R. 2005. Multievent: An extension of multistate capture-recapture models to uncertain states. Biometrics **61(2)**:442–447.

Raudenbush, S., M.-L. Yang, and M. Yosef. 2000. Maximum likelihood for generalized linear models with nested random effects via high-order, multivariate laplace approximation. Journal of Compuational and Graphical Statistics **9(1)**:141–157.

Read, A., P. Drinker, and S. Northridge. 2006. Bycatch of marine mammals in u.s. and global fisheries. Conservation Biology **20**:163–169.

Robbins, J. and D. K. Mattila, 2004. Estimating humpback whale (megaptera nmueungliue) entanglement rates on the basis of scar evidence. final report to the us national marine fisheries service (unpublished). Technical report, Available from the Center for Coastal Studies, Box 1036, Provincetown, MA 02657. 22 pp.

Royle, J. A. 2008. Modeling individual effects in the cormack-jolly-seber model: a state-space formulation. Biometrics **64**:364–370.

Royle, J. A. and R. M. Dorazio. 2008. Hierarchical modeling and inference in ecology: the analysis of data from populations, metapopulations and communities. London, Academic Press Elsevier.

Scott, S. 2002. Bayesian methods for hidden markov models, recursive computing in the 21st century. Journal of the American Statistical Association **97**:337–351.

Seber, G. 1982. The Estimation of Animal Abundance and Related Parameters. 2nd edition. MacMillan Publishing Co., New York, 654pp.

Sethuraman, J. 1994. A constructive definition of dirichlet priors. Statistica Sinica **4**:639–650.

Sharples, R., S. Moss, T. A. Patterson, and P. Hammond. 2012. Spatial variation in foraging behaviour of a marine top predator *(Phoca vitulina)* determined by a large-scale satellite tagging program. PLoS ONE **7(5)**:1–14.

Shirley, K., D. Small, K. Lynch, S. Maisto, and D. Oslin. 2010. Hidden markov models for alcoholism treatment trial data. Annals of Applied Statistics **4**:366–395.

Skaug, H. J. and D. A. Fournier. 2006. Automatic approximation of the marginal likelihood in non-gaussian hierarchical models. Computational Statistics and Data Analysis **51**:699–709.

Spiegelhalter, D., N. Best, B. Carlin, and A. Van Der Linde. 2002. Bayesian measures of model complexity and fit. Journal of the Royal Statistical Society, Series B **64**:583–639.

Stellwagen Bank National Marine Sanctuary, 2006. Shifting the boston traffic separation scheme. national oceanic and atmospheric administration. Http://stellwagen.noaa.gov/science/tss.html.

Stevick, P. T., J. Allen, M. Berube, P. Clapham, S. Katona, F. Larsen, J. Lien, D. Mattila, P. Palsboll, J. Robbins, J. Sigurjonsson, N. Smith, T.and Oien, and P. Hammond. 2003. Segregation of migration by feeding ground origin in north atlantic humpback whales (megaptera novaeangliae). Journal of Zoology **259**:231–237.

Stevick, P. T., J. Allen, P. Chalpham, S. Katona, F. Larsen, J. Lien, D. Mattila, P. Palsboll, R. Sears, J. Sigurjonsson, T. Smith, N. Oien, and P. Hammond. 2006. Population spatial structuring on the feeding grounds in north atlantic humpback whales (megaptera novaeangliae). Journal of Zoology **270(2)**:244–255.

Stevick, P. T., N. Olen, and D. Mattila. 1998. Migration of a humpback whale *(Megaptera Novaeangliae)* between norway and the west indies. Marine Mammal Science **14(1)**:162–166.

Tierney, L. 1994. Markov chains for exploring posterior distributions. The Annals of Statistics **22**:1701–1762.

Tierney, L. and J. Kadane. 1986. Accurate approximation for posterior momnets and marginal densities. Journal of the American Statistical Association **81**:82–86.

Turgeon, J., P. Duchesne, G. Colbeck, L. Postma, and M. Hammill. 2012. Spatiotemporal segregation among summer stocks of beluga *(Delphinapterus leucas)* despite nuclear gene flow: implication for the endangered belugas in eastern hudson bay (canada). Conserv Genet **13**:419–433.

Valenzuela, L., M. Sironi, V. Rowntree, and J. Seger. 2009. Isotopic and genetic evidence for cultrually inherited site fidelity to feeding grounds in southern right whales *(Eubalaena australis)*. Molecular Ecology **18**:782–791.

Weinrich, M., M. Martin, R. Griffiths, J. Bove, and M. Schilling. 1997. A shift in distribution of humpback whales, megaptera novaeangliae, in response to prey in the southern gulf of maine. Fish. Bull. **95**:826–836.

Wilgart, L. 2007. The impacts of anthropogenic ocean noise on cetaceans and implications for management. Can. J. Zool. **85**:1091–1116.

Williams, R., D. Lusseau, and P. Hammond. 2009. The role of social aggregations and protected areas in killer whale conservation: The mixed blessing of ciritcal habitat. Biological Conservation **142**:709–719.

Zucchini, W. and I. MacDonald. 2009. Hidden Markov Models for Time Series An Introduction Using R. Chapman and Hall / CRC.

Zucchini, W., D. Raubenheimer, and I. MacDonald. 2008. Modeling time series of animal behavior by means of a latent-state model with feedback. Biometrics **64(3)**:807–815.